目　　录

第1章　绪　　论

数字图像处理方法的研究源于两个主要应用领域：其一是为了便于人们分析而对图像信息进行改进；其二是为使机器自动理解而对图像数据进行存储、传输及显示。本章有几个主要目的：(1)界定图像处理的范围；(2)回顾一下图像处理领域的历史起源；(3)介绍在图像处理中用以衡量主要应用领域技术性能的概念；(4)简要讨论一下在数字图像处理中进行的主要研究；(5)概述一下典型通用的图像处理系统的组成；(6)列出一些图像处理方面的参考文献。

1.1　数字图像处理的概念

一幅图像可定义为一个二维函数 $f(x,y)$，这里 x 和 y 是空间坐标，而在任何一对空间坐标 (x,y) 上的幅值 f 称为该点图像的强度或灰度。当 x，y 和幅值 f 为有限的、离散的数值时，称该图像为数字图像。数字图像处理是指借用数字计算机处理数字图像，值得提及的是数字图像是由有限的元素组成的，每一个元素都有一个特定的位置和幅值，这些元素称为图像元素、画面元素或像素。像素是广泛用于表示数字图像元素的词汇。在第2章，将用更正式的术语研究这些定义。

视觉是人类最高级的感知器官，所以，毫无疑问图像在人类感知中扮演着最重要的角色。然而，人类感知只限于电磁波谱的视觉波段，成像机器则可覆盖几乎全部电磁波谱，从伽马射线到无线电波。它们可以对非人类习惯的那些图像源进行加工，这些图像源包括超声波、电子显微镜及计算机产生的图像。因此，数字图像处理涉及各种各样的应用领域。

图像处理涉及的范畴或其他相关领域(例如，图像分析和计算机视觉)的界定在初创人之间并没有一致的看法。有时用处理的输入和输出内容都是图像这一特点来界定图像处理的范围。我们认为这一定义仅是人为界定和限制。例如，在这个定义下，甚至最普通的计算一幅图像灰度平均值的工作都不能算做是图像处理。另一方面，有些领域(如计算机视觉)研究的最高目标是用计算机去模拟人类视觉，包括理解和推理并根据视觉输入采取行动等。这一领域本身是人工智能的分支，其目的是模仿人类智能。人工智能领域处在其发展过程中的初期阶段，它的发展比预期的要慢得多，图像分析(也称为图像理解)领域则处在图像处理和计算机视觉两个学科之间。

从图像处理到计算机视觉这个连续的统一体内并没有明确的界线。然而，在这个连续的统一体中可以考虑三种典型的计算处理(即低级、中级和高级处理)来区分其中的各个学科。低级处理涉及初级操作，如降低噪声的图像预处理，对比度增强和图像尖锐化。低级处理是以输入、输出都是图像为特点的处理。中级处理涉及分割(把图像分为不同区域或目标物)以及缩减对目标物的描述，以使其更适合计算机处理及对不同目标的分类(识别)。中级图像处理是以输入为图像，但输出是从这些图像中提取的特征(如边缘、轮廓及不同物体的标识等)为特点的。最后，高级处理涉及在图像分析中被识别物体的总体理解，以及执行与视觉相关的识别函数(处在连续统一体边缘)等。

根据上述讨论,我们看到,图像处理和图像分析两个领域合乎逻辑的重叠区域是图像中特定区域或物体的识别这一领域。这样,在本书中,我们界定数字图像处理包括输入和输出均是图像的处理,同时也包括从图像中提取特征及识别特定物体的处理。举一个简单的文本自动分析方面的例子来具体说明这一概念。在自动分析文本时首先获取一幅包含文本的图像,对该图像进行预处理,提取(分割)字符,然后以适合计算机处理的形式描述这些字符,最后识别这些字符,而所有这些操作都在本书界定的数字图像处理的范围内。理解一页的内容可能要根据理解的复杂度从图像分析或计算机视觉领域考虑问题。这样,本书定义的数字图像处理的概念将在有特殊社会和经济价值的领域内通用。在以下各章展开的概念是那些应用领域所用方法的基础。

1.2　数字图像处理的起源

数字图像处理最早的应用之一是在报纸业,当时,图像第一次通过海底电缆从伦敦传往纽约。早在20世纪20年代曾引入Bartlane电缆图片传输系统,把横跨大西洋传送一幅图片所需的时间从一个多星期减少到3个小时。为了用电缆传输图片,首先要进行编码,然后在接收端用特殊的打印设备重构该图片。图1.1就是用这种方法传送并利用电报打印机通过字符模拟中间色调还原出来的图像。

这些早期数字图像视觉质量的改进工作,涉及到打印过程的选择和亮度等级的分布等问题。用于得到图1.1的打印方法到1921年底就被彻底淘汰了,转而支持一种基于光学还原的技术,该技术在电报接收端用穿孔纸带打出图片。图1.2就是用这种方法得到的图像,对比图1.1,它在色调质量和分辨率方面的改进都很明显。



图1.1　1921年由电报打印机采用特殊字符在编码纸带中产生的数字图像(McFarlane)①

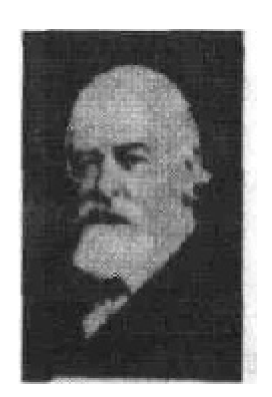


图1.2　1922年在信号两次穿越大西洋后,从穿孔纸带得到的数字图像,可以看出某些差错(McFarlane)

早期的Bartlane系统可以用5个灰度等级对图像编码,到1929年已增加到15个等级。图1.3所示的这种典型类型的图像就是用15级色调设备得到的。在这一时期,由于引入了一种用编码图像纸带去调制光束而使底片感光的系统,明显地改善了复原过程。

刚才引用的数字图像的例子并没有考虑数字图像处理的结果,这主要是因为没有涉及到

① 本书后面提供的参考文献按作者姓氏的字母顺序列出。

计算机。因此,数字图像处理的历史与数字计算机的发展密切相关。事实上,数字图像要求非常大的存储和计算能力,因此数字图像处理领域的发展必须依靠数字计算机及数据存储、显示和传输等相关技术的发展。

计算机的概念可追溯到5 000多年前中国算盘的发明。近两个世纪以来的一些发展也奠定了计算机的基础。然而,现代计算机的基础还要回溯到20世纪40年代由约翰·冯·诺依曼提出的两个重要概念:(1)保存程序和数据的存储器;(2)条件分支。这两个概念是中央处理单元(CPU)的基础。今天,它是计算机的心脏。从冯·诺依曼开始,引发了一系列重要技术进步,使得计算机以强大的功能用于数字图像处理领域。

简单说,这些进步可归纳为如下几点:

(1)1948年贝尔实验室发明了晶体三极管;

(2)20世纪50年代到20世纪60年代高级编程语言(如COBOL和FORTRAN)的开发;

(3)1958年得州仪器公司发明了集成电路(IC);

(4)20世纪60年代早期操作系统的发展;

(5)20世纪70年代Intel公司开发了微处理器(由中央处理单元、存储器和输入输出控制组成的单一芯片);

(6)1981年IBM公司推出了个人计算机;

(7)20世纪70年代出现的大规模集成电路(LI)所引发的元件微小化革命,20世纪80年代出现了VLSI(超大规模集成电路),现在已出现了ULSI。



图1.3 在1929年从伦敦到纽约用15级色调设备通过电缆传送的Generals Pershing和Foch的未经修饰的照片

伴随着这些技术进步,大规模的存储和显示系统也随之发展起来。这两者均是数字图像处理的基础。

第一台可以执行有意义的图像处理任务的大型计算机出现在20世纪60年代早期。数字图像处理技术的诞生可追溯至这一时期这些机器的使用和空间项目的开发,这两大发展把人们的注意力集中到数字图像处理的潜能上。利用计算机技术改善空间探测器发回的图像的工作,始于1964年美国加利福尼亚的喷气推进实验室。当时由“旅行者7号”卫星传送的月球图像由一台计算机进行了处理,以校正航天器上电视摄像机中各种类型的图像畸变。图1.4显示了由“旅行者7号”于1964年7月31日上午(东部白天时间)9点09分在光线影响月球表面前约17分钟时摄取的第一张月球图像[痕迹(称为网状痕迹)用于几何校正,在第5章将讨论该问题],这也是美国航天器取得的第一幅月球图像。“旅行者7号”传送的图像可作为改善的增强和复原图像(例如来自“探索者”登月飞行、“水手号”系列空间探测器及阿波罗载人登月飞行的图像)方法的基础。

进行空间应用的同时,数字图像处理技术在20世纪60年代末和20世纪70年代初开始用于医学图像、地球遥感监测和天文学等领域。早在20世纪70年代发明的计算机轴向断层术(CAT)[简称计算机断层(CT)]是图像处理在医学诊断领域最重要的应用之一。计算机轴向断层术是一种处理方法,在这种处理中,一个检测器环围绕着一个物体(或病人),并且一个X射



图1.4 美国航天器传送的第一张月球照片，"旅行者7号"卫星1964年7月31日9点09分(东部白天时间)在光线影响月球表面前17分钟时摄取的图像

线源(与检测器环同心)绕着物体旋转。X射线穿过物体并由位于对面环中的相应检测器收集起来。当X射线源旋转时，重复这一过程。断层技术由一些算法组成，该算法用感知的数据去重建通过物体的"切片"图像。当物体沿垂直于检测器的方向运动时就产生一系列这样的"切片"，这些切片组成了物体内部的再现图像。断层技术是由Godfrey N. Hounsfield先生和Allan M. Cormack教授发明的，他们共同获得了1979年诺贝尔医学奖。X射线是在1895年由威廉·康拉德·伦琴发现的，由于这一发现，他获得了1901年诺贝尔物理学奖。这两项发明相差近100年。它们在今天引领着图像处理某些最活跃的应用领域。

从20世纪60年代至今，图像处理领域已得到了生机勃勃的发展。除了医学和空间项目的应用外，数字图像处理技术现在已应用在了更广泛的范围。计算机程序用于增强对比度或将亮度编码为彩色，以便于解释X射线和用于工业、医学及生物科学等领域的其他图像。地理学用相同或相似的技术从航空和卫星图像中研究污染模式。图像增强和复原过程用于处理不可修复物体的已损图像或者造价昂贵不可复制的实验结果。在考古学领域，使用图像处理方法已成功地复原了模糊的图片，这些图片是丢失或损坏的稀有物品惟一现存的记录。在物理学和相关领域，计算机技术通常增强如高能等离子和电子显微镜方法等领域的实验图像。图像处理技术也成功地应用在天文学、生物学、核医学、法律实施、国防及工业领域中。

这些例子说明图像处理的结果是用于人类分析。在这一章开始提到，数字图像处理技术的第二类主要应用领域是解决机器感知问题。在这种情况下，注意力集中在以更适合计算机处理的形式从图像中提取信息的过程。通常，这一信息有点类似于人类用于解释一幅图像内容的视觉特性。用于机器感知的信息有统计矩、傅里叶变换系数和多维距离测度。在机器感知中，通常使用图像处理技术的典型问题是自动字符识别、用于生产线及检测的工业机器视觉、军事识别、指纹的自动处理、X射线和血样分类处理、用于天气预报和环境鉴定的航空与卫星图像的机器处理。计算机价格性能比的不断下降和通过万维网及互联网的网络通信带宽的扩展，已为数字图像处理技术的持续发展创造了前所未有的机会。其中一些应用领域将在下面各部分加以说明。

1.3 数字图像处理的应用实例

今天，几乎不存在与数字图像处理无关的技术领域，这里讨论的范围只能涉及其应用领域的一小部分。然而，由于篇幅上的限制，在这一节提供的材料无疑将围绕数字图像处理的广度及重要性展开。这里将介绍一些应用领域，其中的每一个都要使用后续各章要展开讲解的常用数字图像处理技术。在本节中示出的许多图像会在本书后面的多个实例中引用，且这里提供的全部是数字图像。

数字图像处理的应用领域多种多样,所以本书在内容组织上尽量达到该技术应用领域的广度。阐述数字图像处理应用范围最简单的一种方法是根据信息源来分类(如可见光、X射线,等等)。在今天的应用中,最主要的图像源是电磁能谱,其他主要的能源包括声波、超声波和电子(以用于电子显微镜方法的电子束的形式)。建模和可视化应用中的合成图像由计算机产生。这一节将简要讨论一下各类信息源及不同的应用领域如何产生图像。把图像变换为数字图像的方法则在下一章讨论。

建立在电磁波谱辐射基础上的图像是最熟悉的,特别是X射线和可见光谱波段的图像。电磁波可定义为以各种波长传播的正弦波,或者认为是一种粒子流,每个粒子包含一定(一束)能量,每束能量称为一个光子。如果光谱波段根据光子能量进行分组,我们会得到如图1.5所示的伽马射线(最高能量)到无线电波(最低能量)的光谱。如图所示的加底纹的条带表达了这样一个事实,即电磁波谱的各波段间并没有明确的界线,而是由一个波段平滑地过渡到另一个波段。

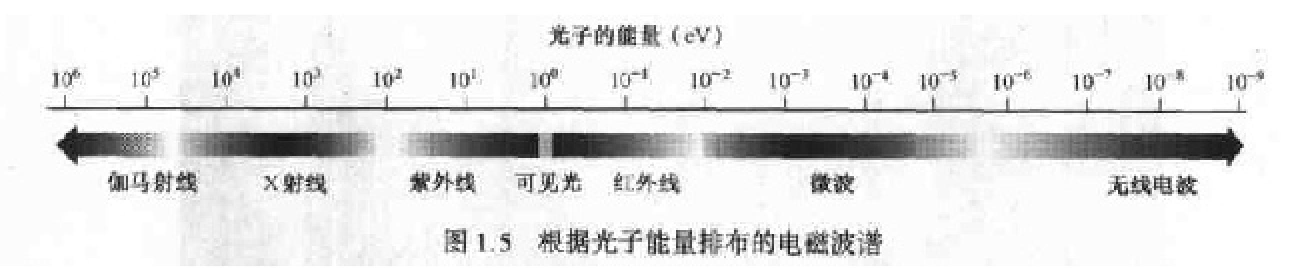


图1.5 根据光子能量排布的电磁波谱

1.3.1 伽马射线成像

伽马射线成像的主要用途包括核医学和天文观测。在核医学中,这种处理是将放射性同位素注射到病人体内,当这种物质衰变时放射出伽马射线,然后用伽马射线检测器收集到的放射物产生图像。图1.6(a)显示了一幅利用伽马射线成像得到的骨骼扫描图像,这类图像用于骨骼病理(例如感染或肿瘤)定位。图1.6(b)显示了另一种叫做"正电子放射断层"(PET)的核成像,其原理与1.2节提到的X射线断层术一样。然而,与使用外部X射线源不同,它给病人注射放射性同位素,同位素衰变时放射出正电子。当正电子遇上一个电子时两者湮没并放射出两束伽马射线。这些射线被检测到后利用断层技术的基本原理创建断层图像。示于图1.6(b)的图像是构成病人三维再现图像序列的一幅样品。这幅图像显示脑部和肺部各有一个肿瘤,即很容易看到的小白块。

大约在15 000年前,天鹅星座中的星星发生大爆炸,产生了一团过热的稳定气云(即天鹅星座环),该气云以彩色阵列形式发光。图1.6(c)显示了在伽马射线波段成像的天鹅星座环。与图1.6(a)和(b)不同,该图像是利用成像物体自然辐射得到的。最后,图1.6(d)显示了一幅来自核反应器电子管的伽马辐射图像,在图像的左下部可以看到较强的辐射区。

1.3.2 X射线成像

X射线是最早用于成像的电磁辐射源之一。最熟悉的X射线应用是医学诊断,但是,X射线还被广泛用于工业和其他领域,像天文学。用X射线管产生用于医学和工业成像的X射线。X射线管是带有阴极和阳极的真空管。阴极加热释放自由电子,这些电子以很高的速度向阳极流动,当电子撞击一个原子核时,能量被释放并形成X射线辐射。X射线的能量由另一

边的阳极电压控制,而 X 射线的数量由施加于阴极灯丝的电流控制。图 1.7(a)显示了一幅位于 X 射线源和对 X 射线能量敏感的胶片之间的病人胸部图像。X 射线的强度受射线穿过病人时的吸收量调制,最终能量落在胶片上并使其感光,这与光使照相胶片感光是一样的。在数字 X 射线照片中,数字图像可用两种方法得到:(1)用数字化的 X 射线胶片;(2)X 射线通过病人身体直接落到某装置上(如荧光屏),该装置把 X 射线转换为光,然后转换而来的光信号由高灵敏度的数字系统获取。第 2 章中将讨论数字化问题。

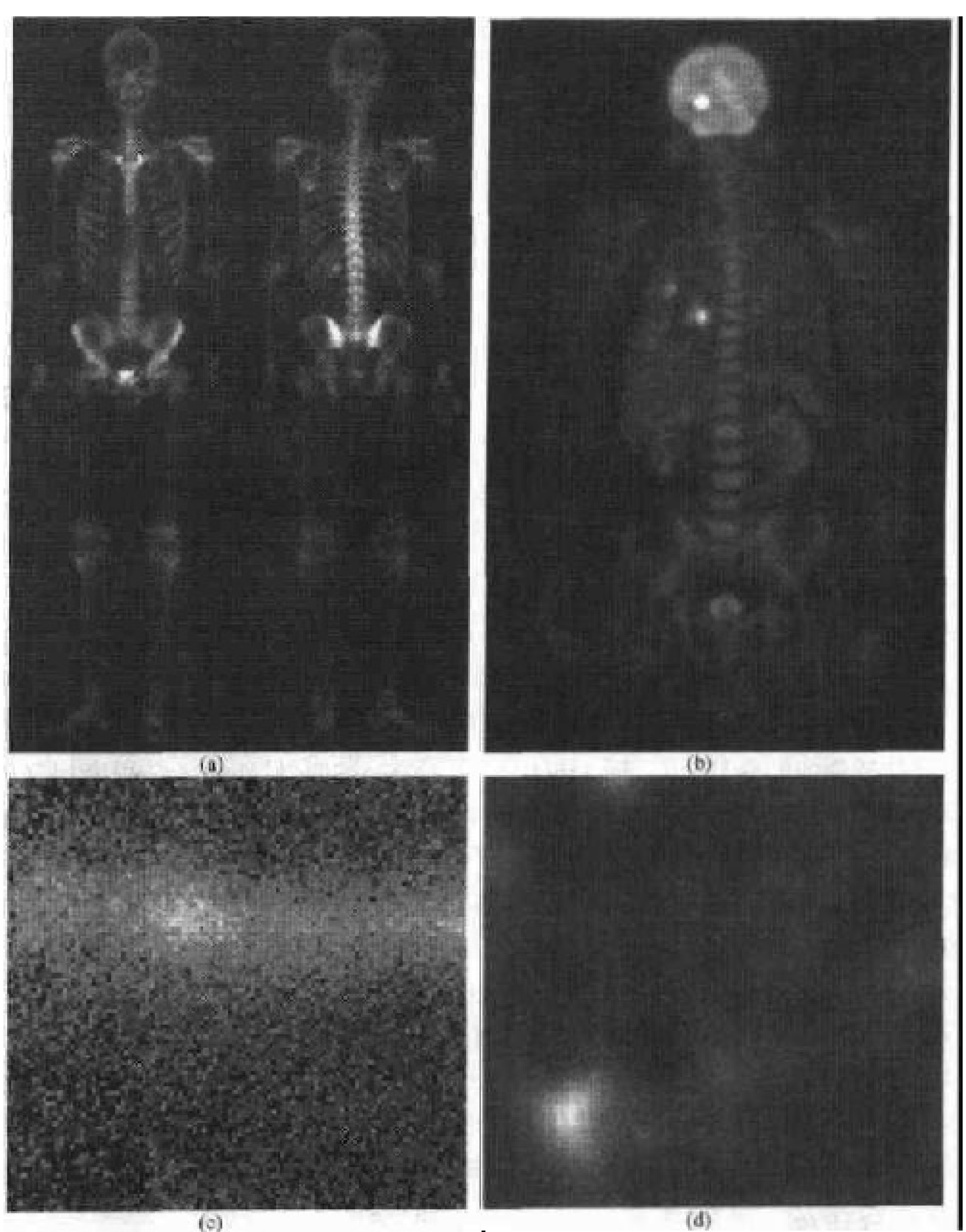


图 1.6 伽马成像实例。(a) 骨骼扫描,(b)PET 图像,(c)天鹅星座环,(d)来自反应器电子管的伽马辐射(亮点)[图像(a)由 GE 医学系统提供,(b)由 CTI PET系统的 Michael E. Casey博士提供,(c)由 NASA提供,(d)由密歇根大学的 Zhong He和 David K. Wehe教授提供]

血管照相术是对比增强辐射成像领域中另一主要应用。这一过程用于得到血

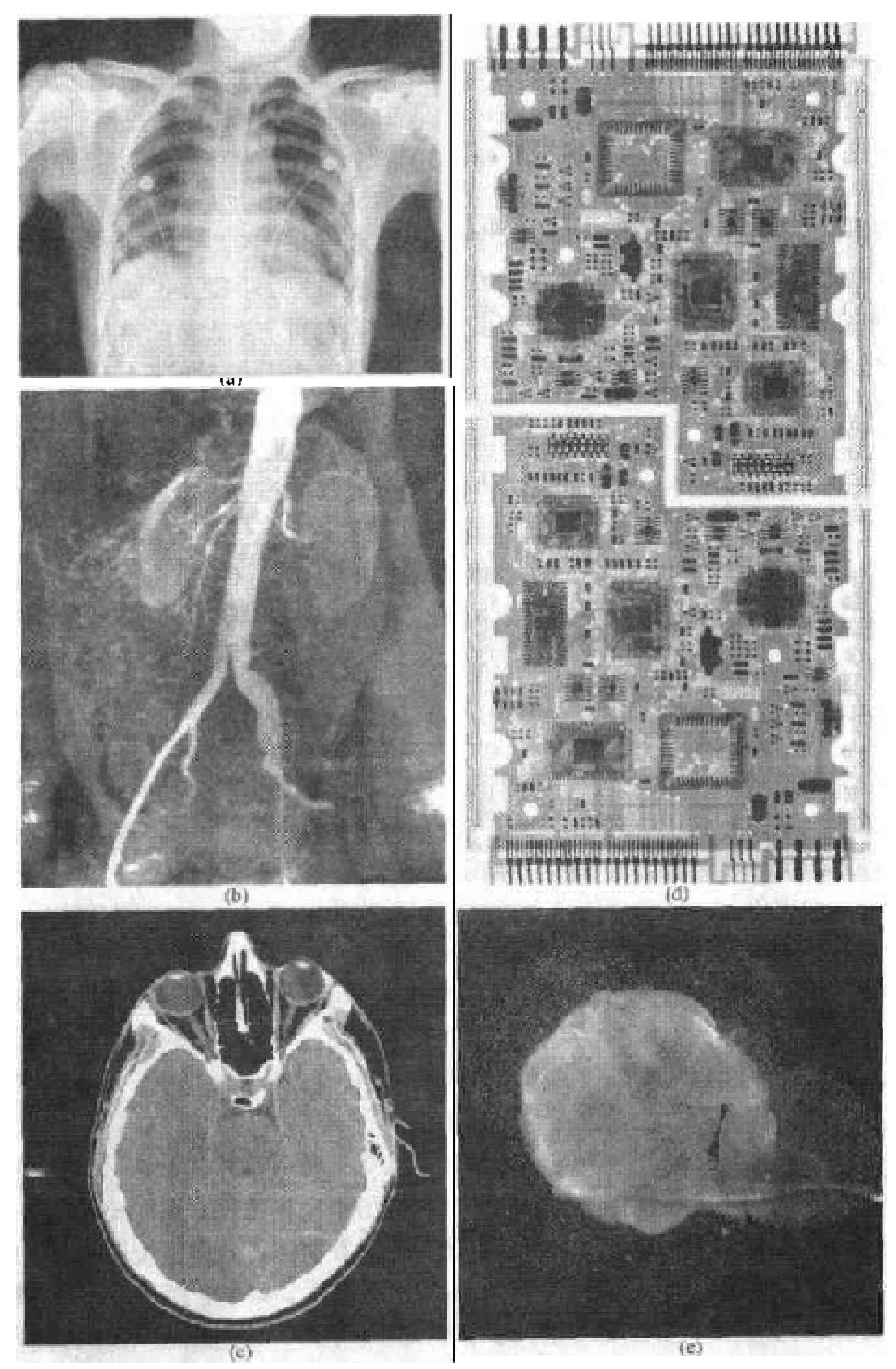


图1.7　X射线成像的实例。(a)胸部X射线图像,(b)主动脉造影图像,(c)头部CT,(d)电路板,(e)天鹅星座环[图像(a)和(c)由Vanderbilt大学医学中心应用辐射学和放射学系的David R. Pickens博士提供,(b)由密歇根大学医学院解剖学分部的Thomas R. Gest博士提供,(d)由Lixi公司的Joseph E. Pascente先生提供,(e)由NASA提供]

管图像(称为血管造影照片)。一个导管(小的、柔软且中空的管子)插入动脉或静脉,导管穿过

血管并被引导到要研究的区域。当导管到达探查的部位时,X 射线对比介质通过导管注入,这就增强了血管的对比并可以观察到任何病变或阻塞。图 1.7(b)显示了一个主动脉血管造影照片的例子。在图像左下角可以看到导管插入血管中。注意在对比介质流向肾时,大血管的强对比效果,这一点在图像中可以看到。正如第 3 章要讨论的,血管照相术是数字图像处理的主要领域,在这里,图像减影法用于增强要研究的血管特性。也许大家最熟悉的 X 革命性的医学处理射线在医学成像中的应用是计算机轴向断层术。由于该技术的分辨率和三维能力,CAT 早在 20 世纪 70 年代第一次付诸使用的时候就改革了医疗手段。正如在 1.2 节提到的,CAT 图像是垂直通过病人的一个"切片",当病人纵向移动时可产生大量的"切片",这些图像组合在一起就构成了病人体内的三维描绘图,其纵向分辨率与切片数量成正比。图 1.7(c)显示了典型的头部 CAT 切片图像。

与刚刚讨论的有些类似的技术也可用于工业处理,但通常涉及较高能量的 X 射线。图 1.7(d)显示了一幅电路板的 X 射线图像。这类图像是 X 射线在工业中两类应用中典型的一例。它们在生产中用于检测电路板的缺陷,如元件缺失或断线等。当部件可以被 X 射线穿透时(例如,塑料组件,甚至更大的物体,像固体发动机和火箭等),CAT 扫描很有用。图 1.7(e)显示了天文学中 X 射线成像的一个例子。这幅图像是图 1.6(c)的天鹅星座环,但此时是 X 射线波段成像。

1.3.3 紫外波段成像

紫外光的应用是多种多样的,包括平板印刷技术、工业检测、显微镜方法、激光、生物图像以及天文观测等。在这一波段中只用显微镜方法和天文观测作为例子来说明这类成像。

紫外光被用于荧光显微镜方法,这是显微镜方法中发展最快的领域之一。荧光是在 19 世纪中叶发现的一种现象,当时,当紫外光直接照射到矿物质上时,第一次发现矿物质荧石发荧光。紫外光本身并不可见,但当紫外辐射光子与荧光材料内原子中的电子碰撞时,它把电子提高到较高的能级,随后,受激电子释放到较低的能级并以可见光范围内的低能光子形式发光。荧光显微镜方法的基本任务是用激发光照射一个样品,然后从较强的激发光中分离出较弱的荧光,这样,仅仅辐射光能到达人眼或其他检测器。结果,荧光区的光对照一个暗背景在某个方向上有足够强的对比度。非荧光材料的背景越暗,设备越有效。

荧光显微镜方法在研究可产生荧光的材料时的确是一种很优秀的方法,无论是自然形态(初级荧光)还是经化学处理后具有荧光性能(二级荧光)都同样适用。图 1.8(a)和(b)显示了使用荧光显微镜方法的典型结果。图 1.8(a)显示了普通谷物的荧光显微镜图像,图 1.8(b)显示了被一种真菌感染了的谷物,谷物、草、葱、高粱等可被 700 多类寄生真菌感染。谷物感染是很严重的危害,因为谷物是世界上最主要的农作物之一。图 1.8(c)显示了在紫外波段高能区成像的天鹅星座环。

1.3.4 可见光及红外波段成像

由于电磁波谱可见光波段在所有波段中是我们最熟悉的,因此,在这一波段的成像应用领域远远超过其他波段就不足为奇了。红外波段常用于与可见光相结合成像,为便于说明,在这一节把可见光和红外光合在一起讨论。在下面的讨论中将涉及光显微镜方法、天文学、遥感、工业和法律实施等方面的应用。

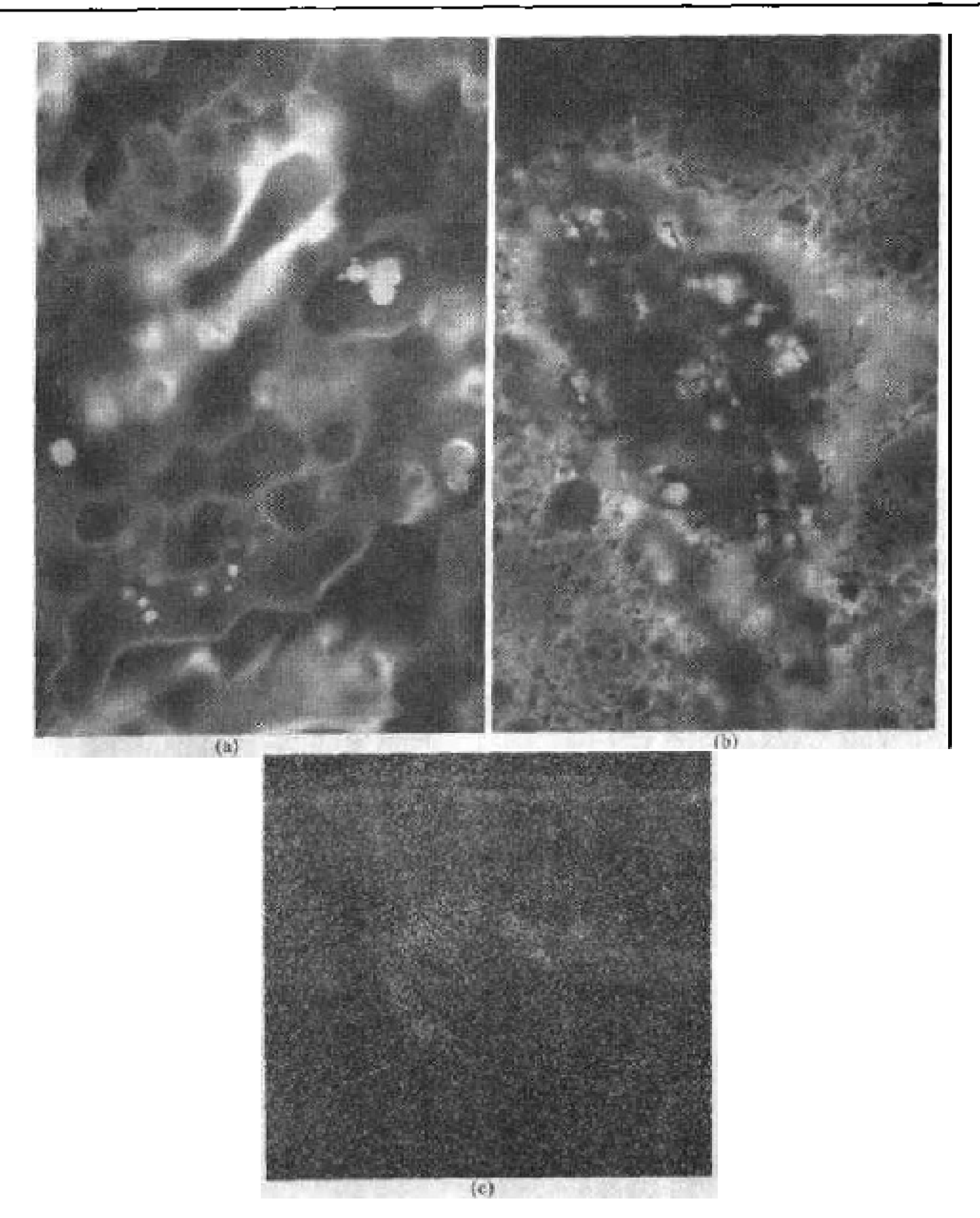


图 1.8 紫外光成像的例子。(a)普通谷物,(b)被"真菌"感染了的谷物,(c) 天鹅星座环[图像(a)和(b)由佛罗里达州立大学的Michael W. Davidson博士提供,(c)由NASA提供]

图 1.9 显示了几个利用光显微镜方法获得的图像的例子。实例涉及的范围从药物到材料特性的显微检测。甚至在显微镜方法中,其应用领域都是很广的,所以不能在这里详细讨论。当然要构思出可能应用于这些图像的处理过程的类型(包括增强和测量等)并不困难。

可见光处理的另一主要应用领域是遥感,通常包括可见光和红外波谱范围的一些波段。表 1.1 显示了在 NASA 的 LANDSAT 卫星中所用的主要波段的性能。LANDSAT 的最初功能是从空间获得并传送地球图像,其目的是监测行星的环境条件。波段用波长来表示,1 μm 等于 10^{-6} m(我们讨论的是电磁波谱的波长范围,更详细的内容见第 2 章),表中标注了每一波段的特性及作用。为了研究这类多光谱图像性能的基本评价,考虑图 1.10,它显示了表 1.1 所列的每一谱段的图像。成像地区是华盛顿,它包括建筑物、道路、植被和通过城市的主要河流等。

人口中心地区的图像常被用来评估人口增长和变迁方式、污染及其他有害环境的因素。可见光和红外光特征之间的区别在这些图像中是十分明显的。例如,在波段 4 和 5 中,从其周围环境可以看出如何较好地界定河流。

表 1.1 NASA 的 LANDSAT 卫星的主要波段

波 段 号	波 段	波 长(μm)	特性和应用
1	可见蓝光	0.45 ~ 0.52	对水有最大的穿透
2	可见绿光	0.52 ~ 0.60	对测量植物生活力较适用
3	可见红光	0.63 ~ 0.69	植被鉴别
4	近红外光	0.76 ~ 0.90	生物团和海岸线测绘
5	中红外光	1.55 ~ 1.75	植被和土壤含水量
6	热红外光	10.4 ~ 12.5	土壤温度、热量测绘
7	中红外光	2.08 ~ 2.35	矿物测绘

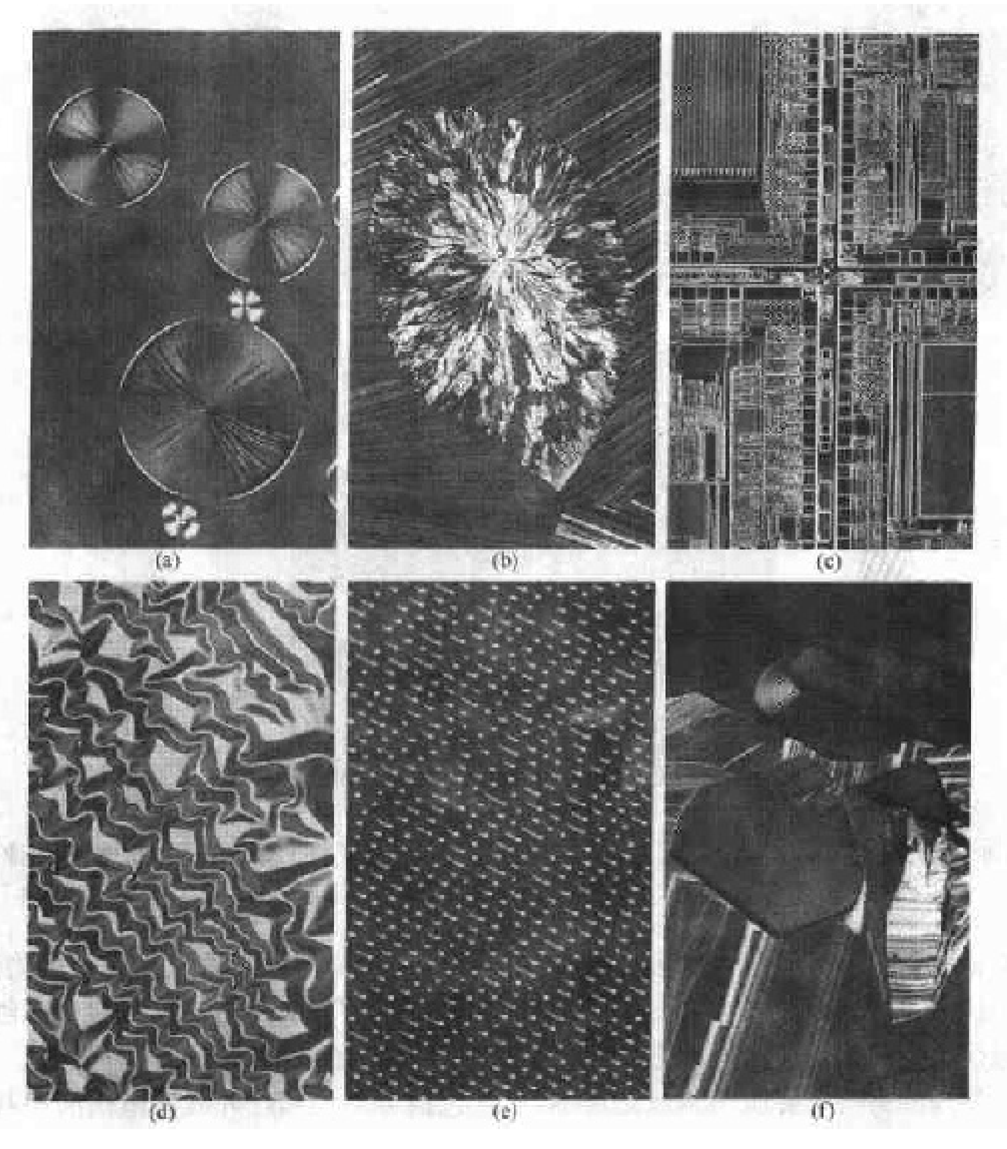


图 1.9 利用可见光显微镜方法获得的图像实例。(a)紫杉酚(抗癌剂),放大了250倍,(b)胆固醇,40倍,(c)微处理器,60倍,(d)镍氧化物薄片,600倍,(e)音频CD的表面,1 750倍,(f)有机超导,450倍(图像由佛罗里达州立大学的Michael W. Davidson博士提供)

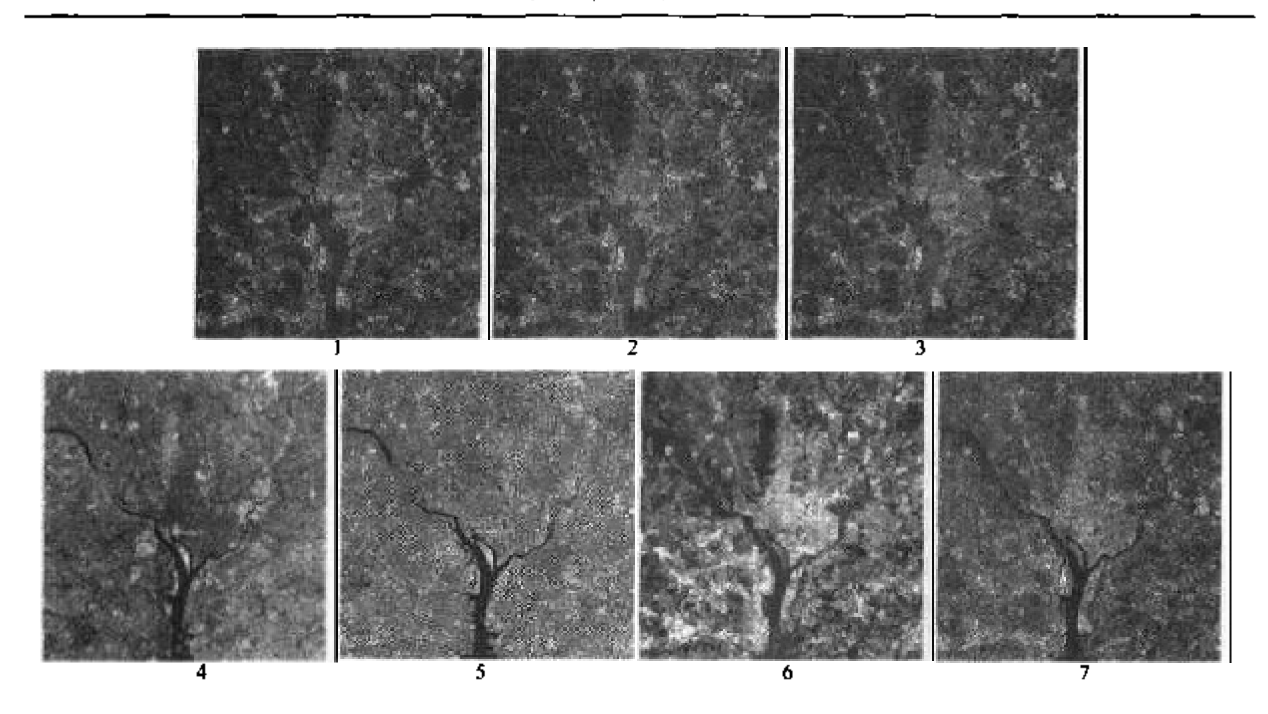


图 1.10　华盛顿区域的 LANDSAT 卫星图像。数字与表 1.1 的波段号对应(图像由 NASA 提供)

天气观测与预报也是卫星多光谱图像的主要应用领域。例如,图 1.11 是由海洋大气管理局的卫星用可见光及红外光传感器取得的飓风图像,在这幅图像中飓风眼清晰可见。

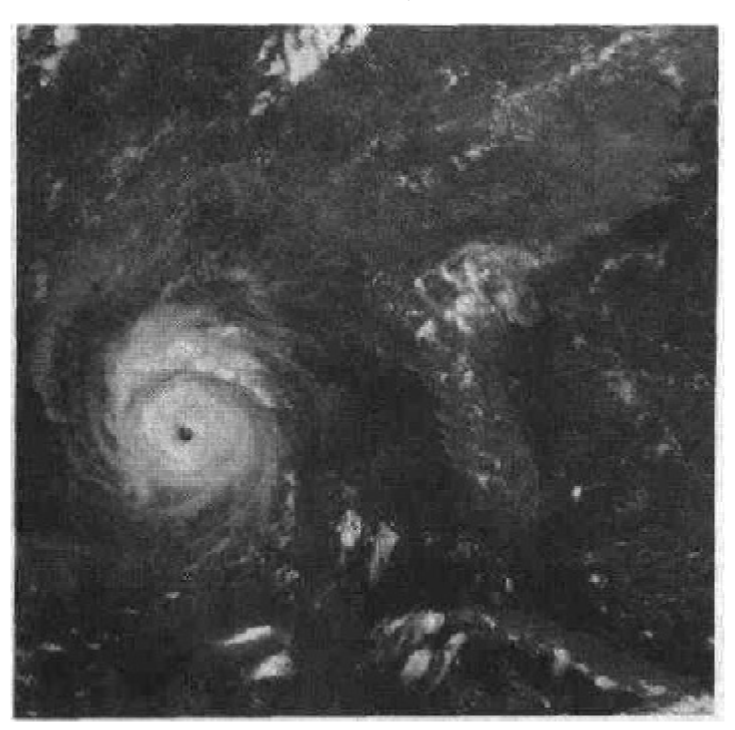


图 1.11　由 NOAA GEOS 传感器获取的安德鲁飓风的多光谱图像(由 NOAA 提供)

图 1.12 和图 1.13 显示了红外图像的应用。这些图像是“全世界夜间灯光”数据集的一部分,它可以提供全球人类居住区的汇总情况。图像是由装在 NOAA DMSP(防御气象卫星项目)卫星上的红外图像系统产生的。红外成像系统在10.0 ~ 13.4 μm波段工作,它具有独特的

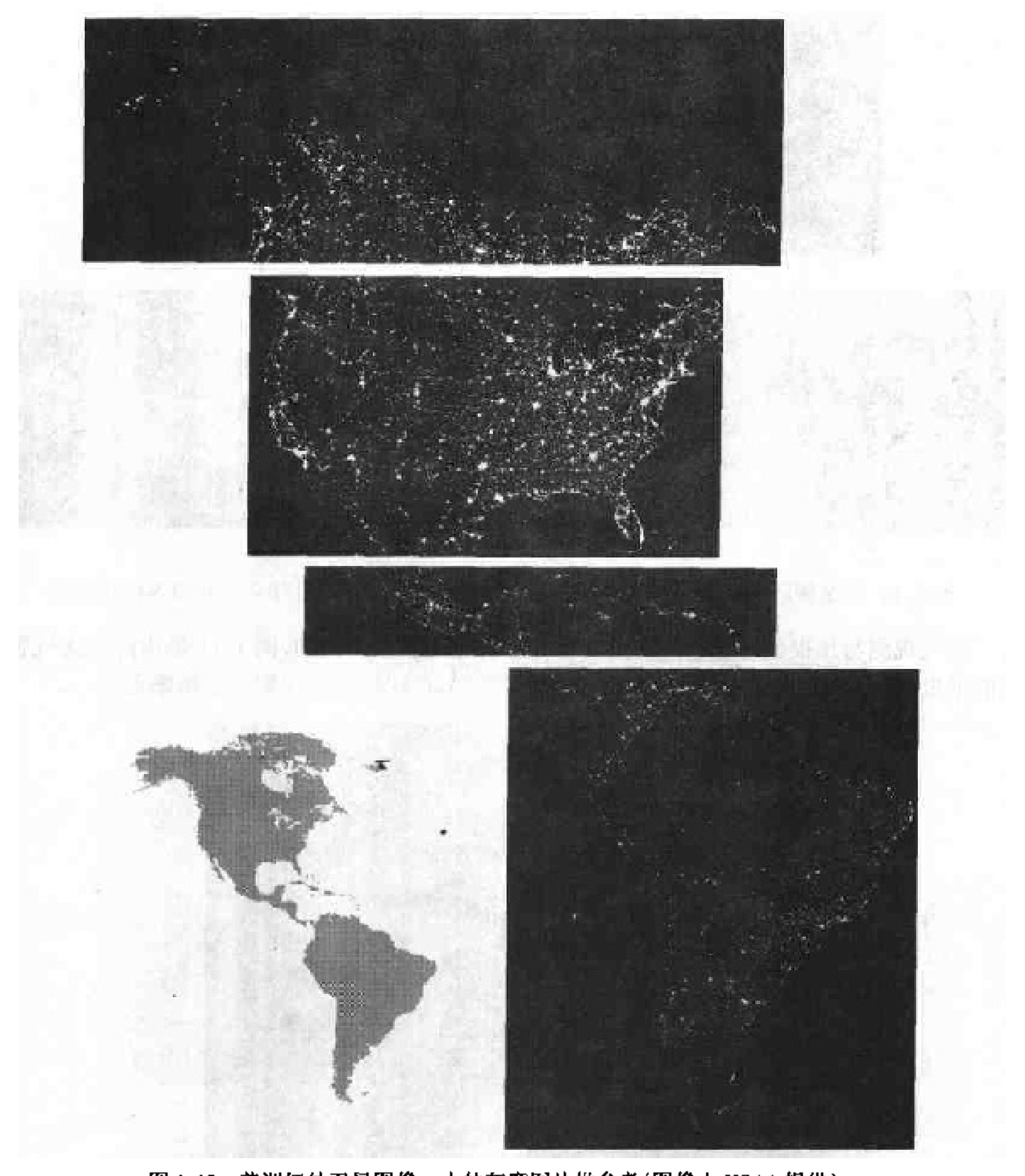


图 1.12 美洲红外卫星图像。小的灰度图片供参考(图像由 NOAA 提供)

功能去观察地球表面微弱的接近可见光的近红外光发射源,包括城市、小镇、村庄、气体火焰及火光。即使在图像处理中没有正式地训练过,编写一个计算机程序,用这些图像去估计世界不同地区使用的电能总数的百分比也是不困难的。在可见光谱中主要成像领域是生产产品的自动视觉检测。图 1.14 显示了一些例子。图 1.14(a)是一块 CD-ROM 装置控制板,一个典型的图像处理任务是检测丢失的部件(在图像右上角 1/4 处是一个缺少部件的例子)。图 1.14(b)是丸剂胶囊的图像,这里的目标是用机器寻找缺少的药丸。图 1.14(c)显示了一种应用,在该应用中,图像处理工作用来寻找没有装到要求水量的瓶子。图 1.14(d)显示了一个干净的塑

料部件,部件内带有不可接受的数量的气泡。像这些异常检测是工业检测的主旨,这类检测也涉及其他产品,如木材和布匹。图1.14(e)显示了一组谷物,通过颜色和存在的异常(如变黑的叶子)等进行检测。最后图1.14(f)显示了一幅有掺杂物的目镜图像(位于人眼处的镜头)。使用一种"结构光"照明技术用强光照射镜头中心很容易检测镜头畸变。在类似钟表表盘1点钟和5点钟处的刻痕是由钳子造成的损坏,其他的多数小斑点是残片。这类检测目标是在封装期间自动地找出缺损和不正确的加工装配。

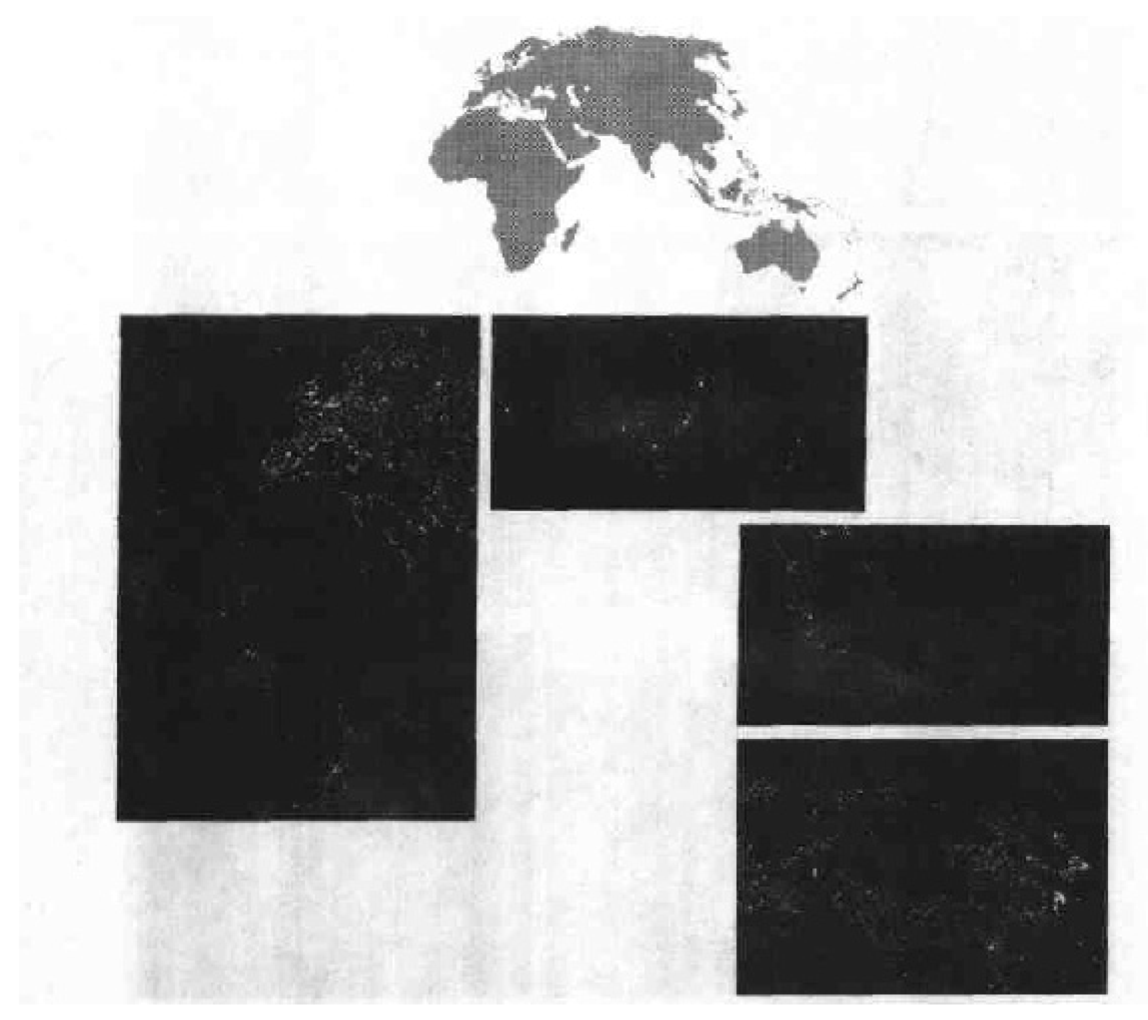


图1.13 世界常住人口聚居区的红外卫星图像,小的灰度图片供参考(由NOAA提供)

作为可见光谱中图像处理的最后说明,可考虑图1.15。图1.15(a)显示了一幅拇指指纹图像。指纹图像常由计算机来处理,该处理不是增强指纹图像就是寻找特征,以便自动搜索数据库寻找潜在的匹配。图1.15(b)显示了一幅纸币图像。数字图像处理在这个领域的应用包括自动计数、法律实施、为跟踪和鉴别钱币读取系列号等。两幅汽车图像如图15(c)和(d)所示,它们是读取牌照的例子。亮的方块指出一个区域,在该区域内图像系统检测该平面,黑的方块显示系统自动读取内容的结果。牌照和其他字符识别的应用广泛用于交通监测和监视中。

1.3.5 微波波段成像

微波波段成像的典型应用是雷达。成像雷达的独特之处是在任何范围、任何时间内,不考

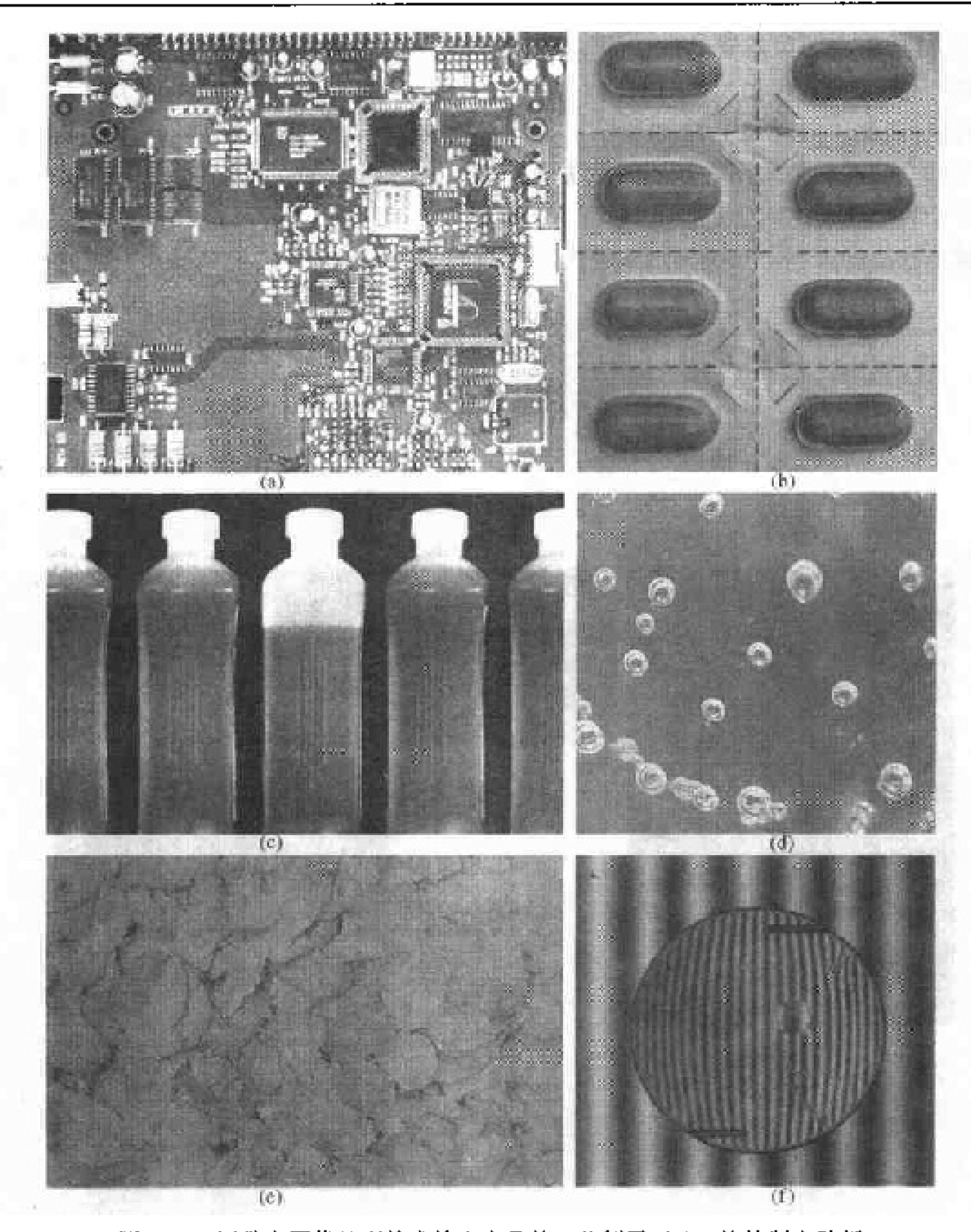


图 1.14 用数字图像处理技术检查产品的一些例子。(a)一块控制电路板,(b)封装的丸剂,(c)瓶子,(d)清洁塑料上的气泡,(e)谷物,(f)目镜搀杂物图像[图(f)由Perceptics公司的Pete Sites先生提供]

虑气候、周围光照条件的收集数据的能力。雷达波可以穿透云层,在一定条件下还可以通过植被、冰层和极干燥的沙漠。在许多情况下,雷达是探测地球表面不可接近地区的惟一方法。成像雷达的工作原理像一个闪光照相机,它自己提供照明(微波脉冲)去照亮地面的一个地区,并拍摄快拍图像。与照相机镜头不同,雷达使用天线和数字计算机去记录图像。在雷达图像中,能看到的只是反射到雷达天线的微波能量。

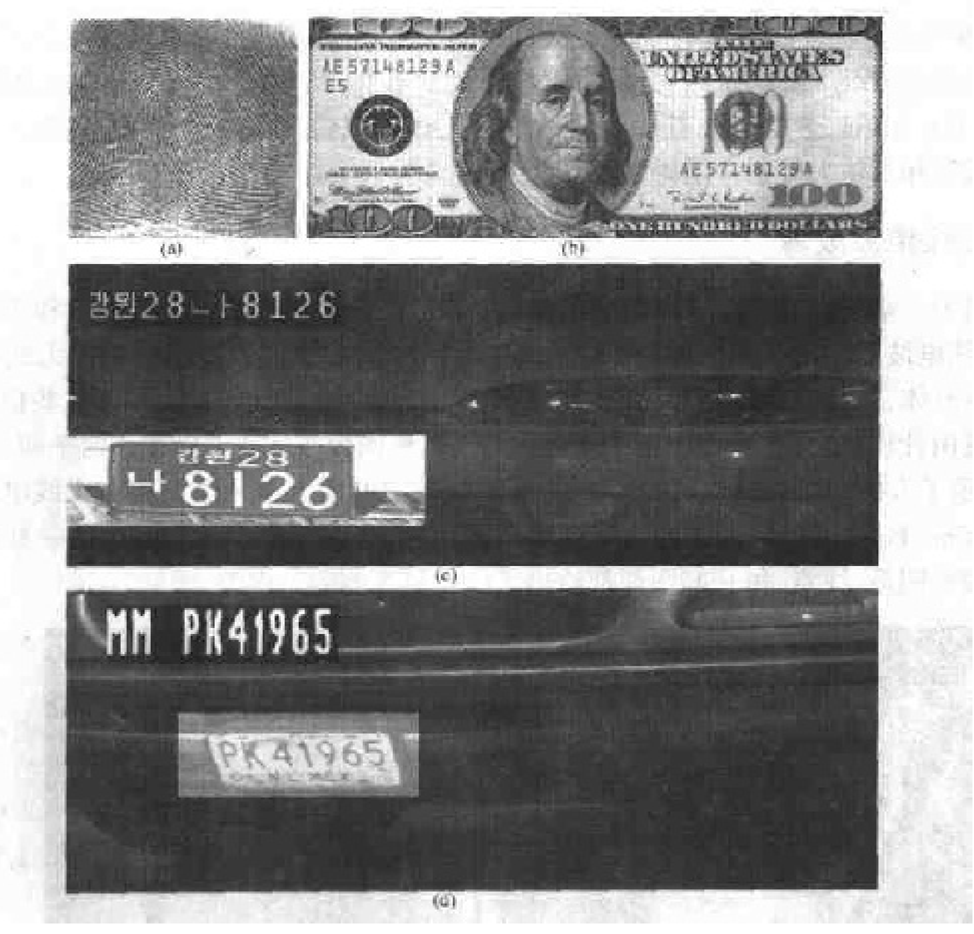


图 1.15　一些可见光谱成像的例子。(a)拇指指纹,(b)纸币,(c)和(d)自动读取牌照[图(a)由美国图像标准技术研究所提供,(c)和(d)由Perceptics公司的Juan Herrera先生提供]

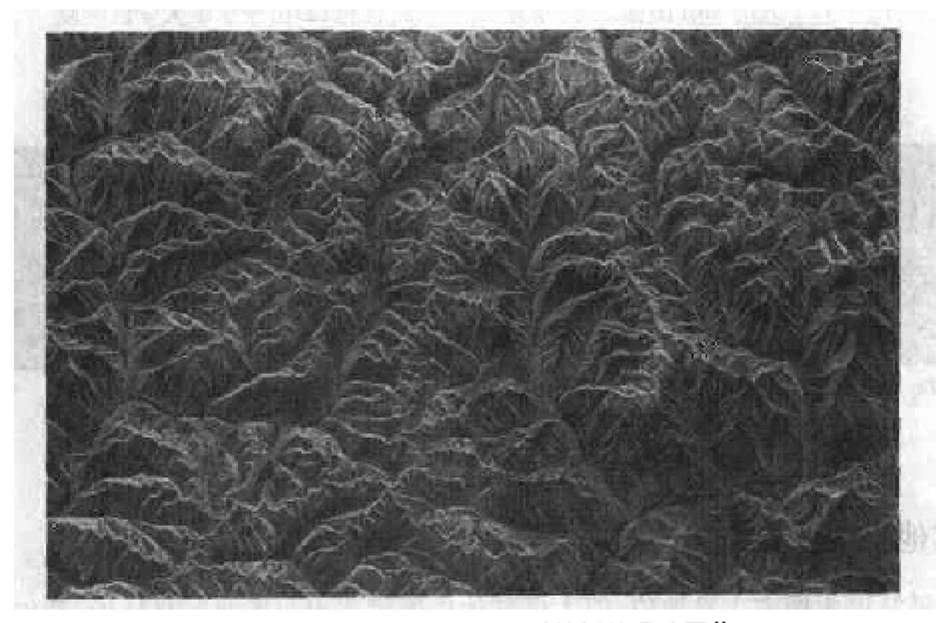


图 1.16　西藏东南山区航天器拍摄的雷达图像

图 1.16 显示了一幅西藏东南方高低不平山区的航天器图像,拉萨市约在东边90 km处,右下角是广阔的拉萨河谷,这是西藏农民和牧民们居住的地方。这一地区的山高约为海拔 5 800 m(19 000 ft,1m = 3.28 ft),其谷底位于海拔 4 300 m(14 000 ft)处。注意,图像的清晰度及细节没有被云和其他大气条件阻碍,通常这些在视觉波段会影响图像。

1.3.6 无线电波成像

如波谱另一端(伽马射线)成像情况那样,无线电波段成像主要应用在医学和天文学。在医学中,无线电波用于磁共振成像(MRI)。该技术是把病人放在强磁场中并使无线电短脉冲通过病人的身体,每一个脉冲将导致一个由病人组织发射的无线电响应脉冲,这些信号发生的位置和强度由计算机确定,从而产生一个病人的横截面图像。MRI 可以在任何平面产生图像。图 1.17 显示了人的膝盖和颈部的图像。图 1.18 最右边的图像显示了一幅在无线电波段的蟹状脉冲星(Crab Pulsar)图像,为了做一个有意义的比较,还显示了用前面讨论的一些波段摄取的相同区域的图像,注意,每一幅图像都给出了脉冲星观测的总误差。

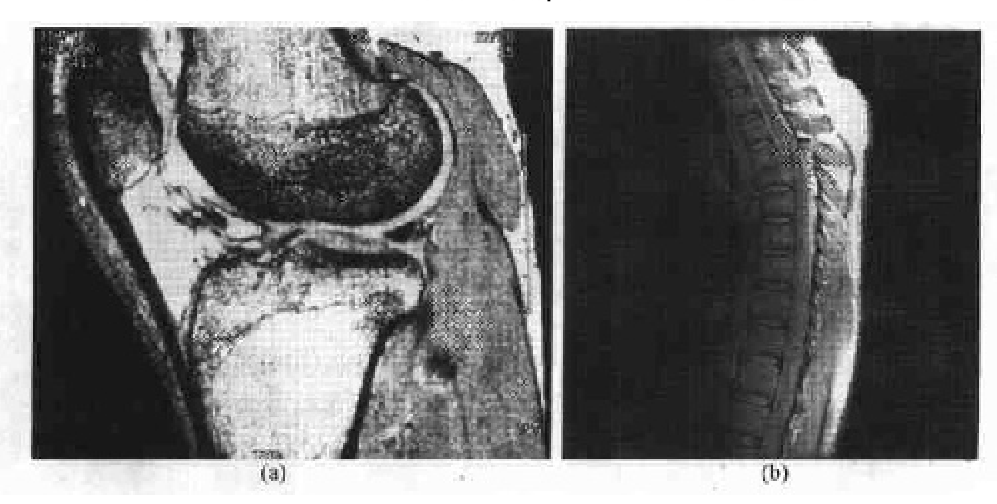


图 1.17 人的 MRI 图像。(a)膝盖,(b)脊锥[图像(a)由密歇根大学医学院解剖科学分部的Thomas R. Gest博士提供,(b)由 Vanderbilt 大学医学中心放射学和放射科学系David R. Pickens博士提供]

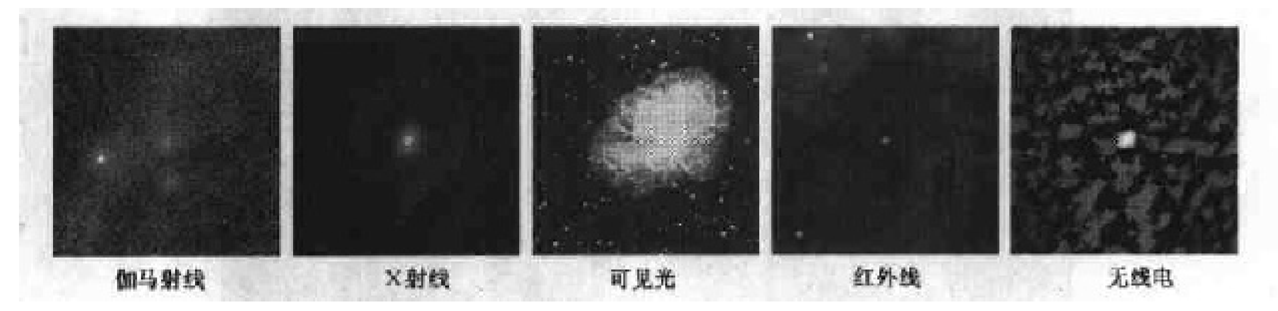


图 1.18 覆盖电磁波谱的蟹状脉冲星图像(位于图像中心)(由 NASA 提供)

1.3.7 其他图像模式应用的实例

虽然电磁波谱成像占主导地位,但大量其他成像模式也很重要。具体地,我们在这一节讨论一下声波成像、电子显微镜方法和合成成像(计算机产生的)。

利用"声音"的成像在地质勘探、工业和医学中得到了应用。地质应用是采用声谱中低端的声波(几百赫兹),在其他应用领域的成像使用超声波(百万赫兹)。图像处理在地质中更重要的商业应用是矿藏和石油勘探。为了透过地表获取图像,主要方法之一是利用一辆矿车和一个大钢制平台,平台由矿车压在地上,并且矿车用频谱为 100 Hz 的频率振动,返回声波的强度和速度由地表下的成分决定,经过计算机分析并由分析结果产生图像。

为了从海洋获取图像,能源通常由两个位于船后部的空气枪组成,返回的声波由置于船后部电缆中的水听器检测。水听器放置在海底,或者用浮子悬吊着(垂直电缆)。两个空气枪交替加压至 2 000 psi(每平方英寸磅数),然后爆炸,船的恒定运动提供了一个横向运动,通过返回声波产生一幅海洋底部地下结构的三维图像。

图 1.19 显示了一幅众所周知的三维模型的横截面图像,与此图像对照,测试了地震成像算法的性能。箭头指向碳氢化合物(油或气)收集器。这一目标比周围明亮,因为目标区域内在亮度方面变化较大。地震解释器寻找那些亮点以便发现油和气。上面层次也是亮的,但是它们的亮度不像横向层亮度那样强。许多地震重建算法由于以上的缺点使对该目标成像较为困难。

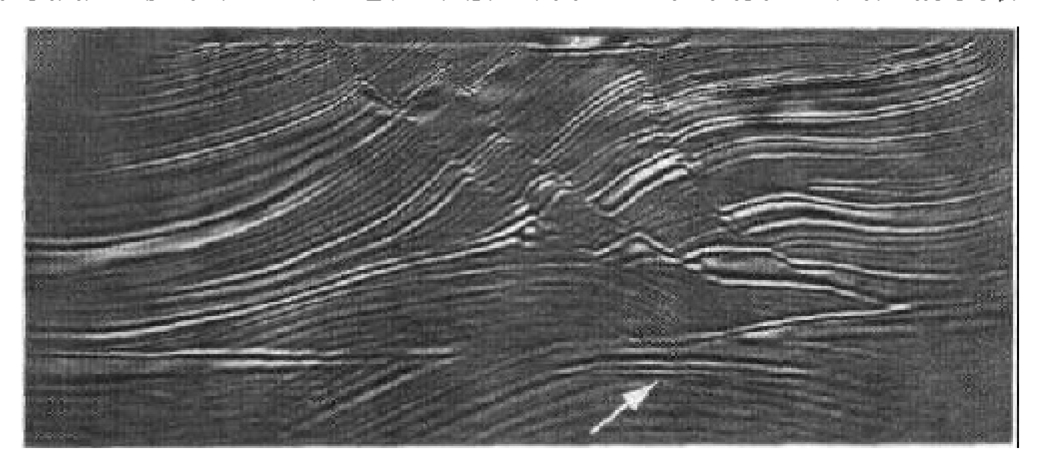


图 1.19　地震模型横截面图像,箭头指示碳氢化合物(由美国 Sandia 国家实验室 Curtis Ober 博士提供)

虽然超声成像常用于制造业,但这一技术最为熟悉的应用是医学,特别是妇产科。在这里,未出生的胎儿被成像以确定他们发育的健康状况。这一检测的副业是确定胎儿的性别。超声波图像是利用下面的基本步骤产生的:

1. 超声波系统(一台计算机、由超声波和接收器组成的超声探头和显示器)向身体传输高频(1~5 MHz)声脉冲。
2. 声波传入体内并碰撞组织间的边缘(流体和软组织之间,软组织和骨骼之间),声波的一部分返回到探头,一部分声波继续传播直到它们到达另一边界并被反射。
3. 反射波被探头收集起来并传给计算机。
4. 计算机根据声波在组织中的传播速度和每个回波返回的时间计算从探头到组织或者器官边界的距离。
5. 系统在屏幕上显示回波的距离和亮度形成的二维图像。

在典型的超声波图像中,每秒钟有上百万个脉冲与回波发出和收集。探头可沿着身体表面以不同角度运动,以便得到各种观察图像。图 1.20 显示了几个例子。

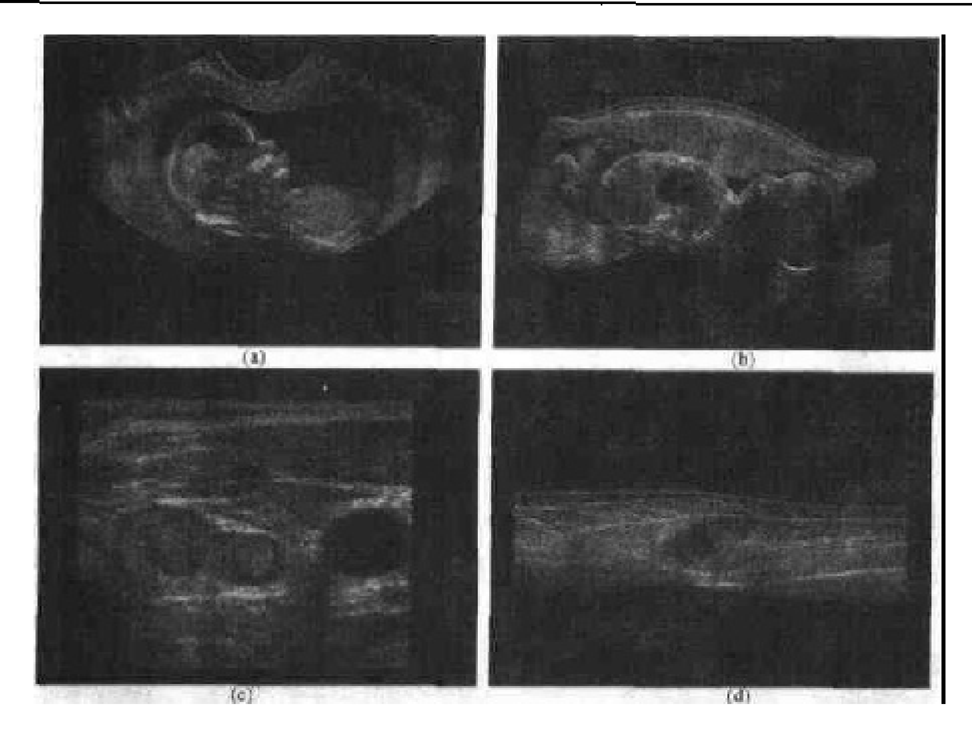


图 1.20 超声波成像的例子。(a)胎儿,(b)胎儿的另一个侧面,(c)甲状腺,(d)肌肉层有损害(由西门子医疗系统有限公司的超声波组织提供)

我们继续用电子显微镜方法的某些例子就成像模式进行讨论。电子显微功能就像它们的镜像,只是它们使用一个聚焦的电子波束代替光束形成图像镜像。电子显微镜的操作包括下面几个基本步骤:由电子源产生一个电子流,用正电势朝着标本加速。该电子束用金属小孔和电磁透镜限制并聚焦为一个薄片形,再聚焦成单一波束,这一波束用磁透镜聚焦到标本上。在标本内部产生照射的交互动作并改变电子束。这些交互动作和影响被检测并转换为图像。这样以相同的方式光被物体反射或吸收。所有的电子显微镜均毫无例外地执行这些基本步骤。

透射(发射)电子显微镜(TEM)的工作很像一个幻灯片投影器。投影器通过幻灯片发射一束光,当光通过幻灯片时,受到幻灯片的内容影响。这一发射光束被投射到观察屏上,形成幻灯片的一个放大了的图像。除了它们发射一个电子束通过标本之外,TEM 以相似的方式工作。通过标本发射的一部分电子流被投影到荧光屏上,伴随着荧光物质和电子的交互作用产生光,由此,得到了可观察的图像。另外,扫描电子显微镜(SEM)实际上扫描电子束并记录每一位置上电子束与样本的交互作用,在荧光屏上产生一个点。一幅完整的图像由通过样本的射束光栅扫描形成,这非常像电视摄像机的原理,电子与荧光屏交互作用并产生光。SEM 适合于大块样品,而 TEM 要求非常薄的样品。

电子显微镜有非常高的放大能力,光学显微镜的放大倍数限制在 1 000 左右。电子显微镜的放大倍数可达到 10 000 倍或更大。图 1.21 显示了由于热过载造成损坏的图像。

我们通过不是从物理目标得到的图像的简单浏览结束成像模式的讨论。这些图像都是用计算机产生的。分形是计算机产生的图像的一个很有影响的例子(Lu [1977])。基本上,分形只不过是根据某些数学定律对一个基本模式的迭代复制。例如,“瓦片”(tiling)是产生分形图像最简单的方法之一。一个方块可以被分成 4 个小方块,这些方块的每一个又可进一步分成 4个较小的方块区,等等。为了填充每一个子方块,根据规则的复杂性,一些瓦片图像可用这

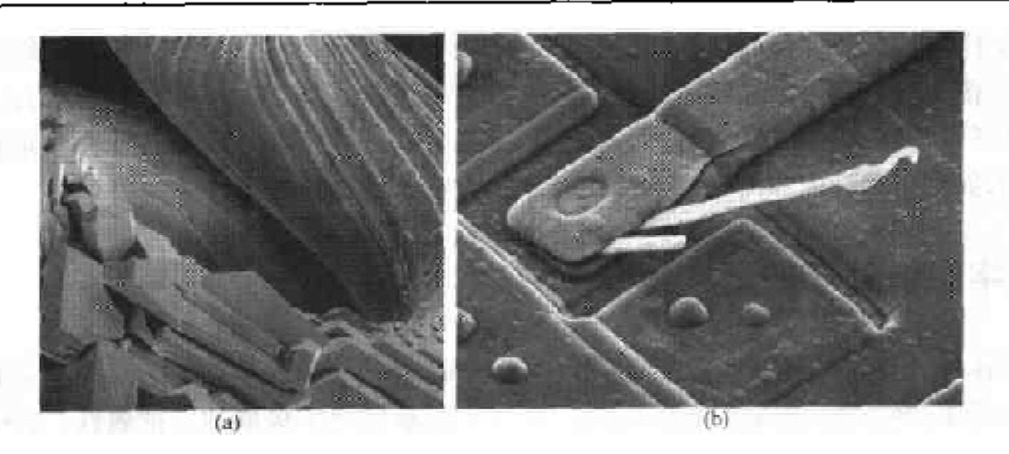


图1.21 (a)过热损坏的钨丝的250倍SEM图像，(b)损坏的集成电路的2 500倍SEM图像。白色纤线是由热破坏造成氧化的结果[图(a)由俄勒冈大学地理科学系的Michael Shaffer先生提供，(b)由加拿大McMaster大学的J.M.Hudak博士提供]

一方法产生。当然，几何形状可以是任意的。例如，分形图像可以从中心点向外辐射生长。图1.22(a)显示了用这种方法生成的分形图像。图1.22(b)显示了另一幅分形图像，它对前述各节例子所用的空间图像提供了一个有趣的相似性。

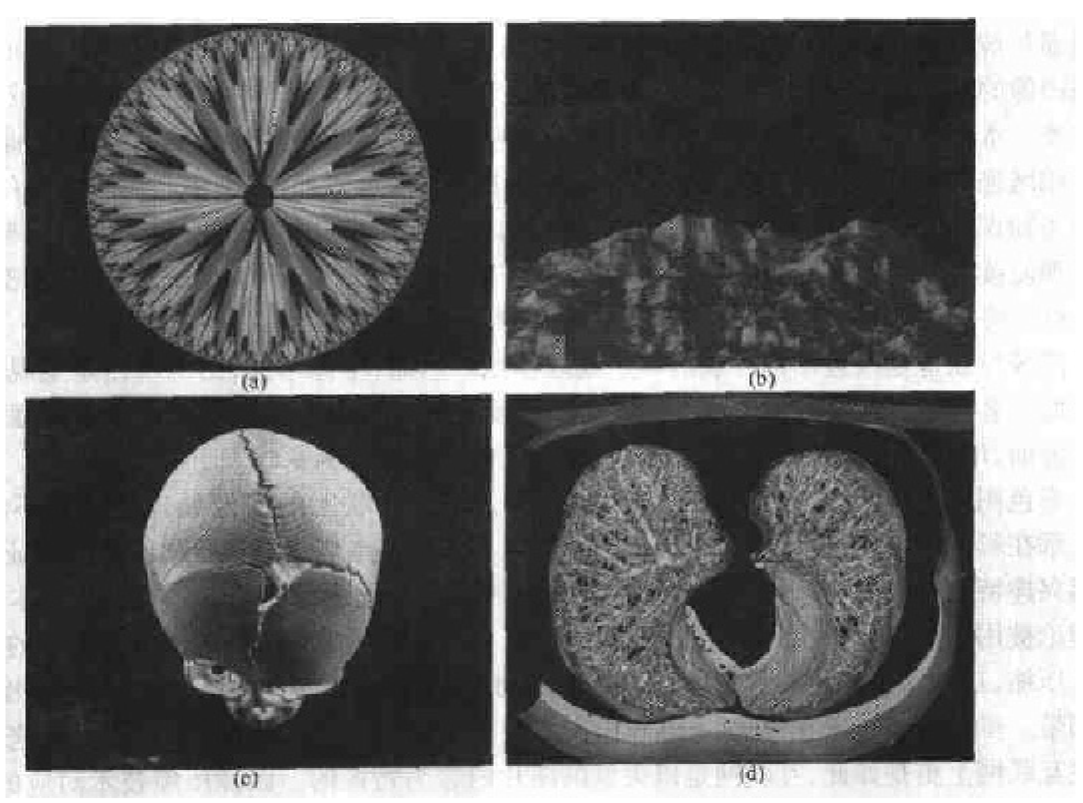


图1.22 (a)和(b)分形图像，(c)和(d)从显示物体的三维计算机模型产生的图像[图(a)和(b)由Swarthmore学院的Melissa D.Binde先生提供，图(c)和(d)由NASA提供]

分形图像朝着艺术的、根据某一规则的子图像元素生长的数学表示方向发展。它们有时对生成随机纹理很有用。借助于计算机更结构化的图像生成研究依赖于三维模型。这是一个在

图像处理和计算机图形学之间提供了重要交叉的领域,也是许多三维可视化系统(如光模拟器)的基础。图 1.22(c)和(d)显示了计算机生成的图像例子。因为原始物体以三维方式产生,图像可以从三维物体的平面投影以任何透视角度产生。这些类型的图像可用于医学培训及其他主要应用,如刑事法庭中和需要产生特殊效果的场合。

1.4 数字图像处理的基本步骤

把后续各章涉及的材料划分为在 1.1 节定义的两个主要类别是有帮助的。一类是其输入和输出都是图像;另一类,输入可能是图像,但是输出是从图像中提取的特征属性。这一结构在图 1.23 中做了概括。该图并不意味着每一步处理都用于图像。更确切的目的是给出所有方法的一个概念,这些方法可以针对不同的目的,也许带有不同的目标。这一节的讨论也可以看做是本书剩余部分材料的综述。

图像获取是图 1.23 所示的第一步处理。在 1.3 节的讨论中给出了关于数字图像来源的提示。这个题目在第 2 章将更具体地进行研究。在这一章,我们还要介绍一些基本的数字图像概念,这些概念将贯穿全书。注意到获取与给出一幅数字形式的图像一样简单。通常,图像获取包括如设置比例尺等预处理。

图像增强是数字图像处理最简单和最有吸引力的领域。基本上,增强技术后面的思路是显现那些被模糊了的细节,或简单地突出一幅图像中感兴趣的特征。一个图像增强的例子是增强图像的对比度,使其看起来好一些。应记住,增强是图像处理中非常主观的领域,这一点很重要。本书有两章涉及增强处理并不是因为它比本书涉及的其他题目更重要,而是因为我们要用增强处理引导读者更好地运用其他章节中的技术。另外,我们不是用一章专门讲述数学预备知识,而是通过说明在图像增强中用到的数学知识来介绍一些所需要的数学概念。这样处理使读者能在运用图像处理技术时熟悉这些概念。一个较好的例子就是傅里叶变换。在第 4 章中将介绍傅里叶变换,但是在其他章节中也有很多应用。

图像复原也是改进图像外貌的一个处理领域。然而,不像增强,图像增强是主观的,而图像复原是客观的。在某种意义上说,复原技术倾向于以图像退化的数学或概率模型为基础。另一方面,增强以怎样构成好的增强效果这种人的主观偏爱为基础。

彩色图像处理已经成为一个重要领域,因为基于互联网的图像处理应用在不断增长。第 5 章在彩色模型、数字域的彩色处理方面涵盖了大量基本概念。在后续各章,彩色还是图像中感兴趣特征被提取的基础。小波是在各种分辨率下描述图像的基础。特别是在本书中,这些理论被用于图像数据压缩及金字塔描述方法。在这里,图像被成功地细分为较小的区域。

压缩,正如其名称所指的意思,所涉及的技术是减少图像的存储量,或者在传输图像时降低频带。虽然存储技术在过去的 10 年内有了很大改进,但对传输能力我们还不能这样说,尤其在互联网上更是如此,互联网是以大量的图片内容为特征的。图像压缩技术对应的图像文件扩展名对大多数计算机用户是很熟悉的(也许没注意),如 JPG 文件扩展名用于 JPEG(联合图片专家组)图像压缩标准。

形态学处理涉及提取图像元素的工具,它在表现和描述形状方面非常有用。正如 1.1 节指出的,这一章的材料将从输出图像处理到输出图像特征处理的转换开始。

分割过程将一幅图像划分为组成部分或目标物。通常,自主分割是数字图像处理中最为

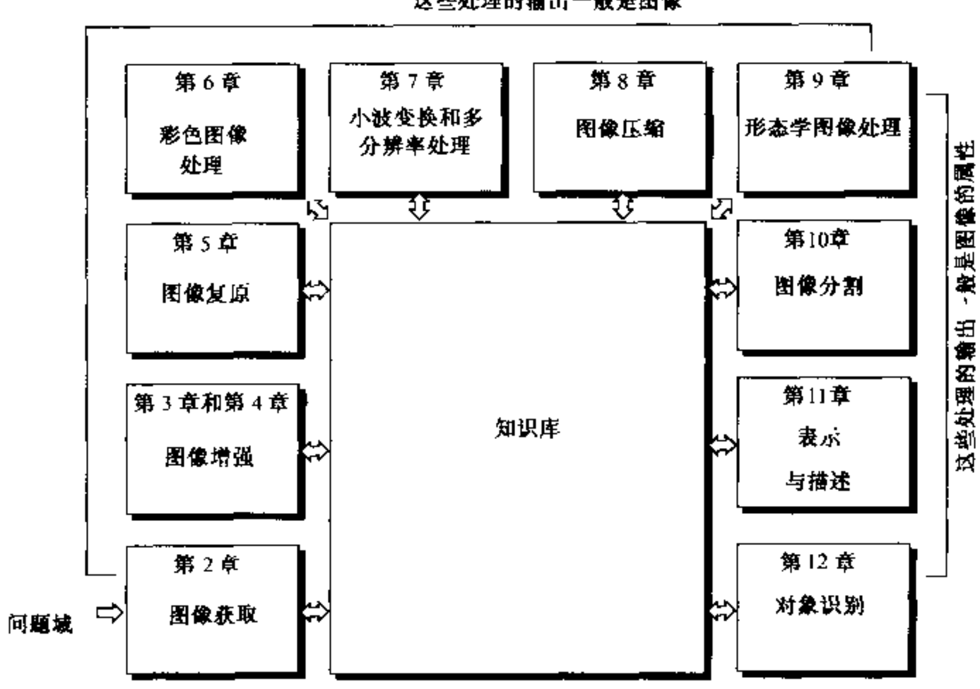


图1.23　数字图像处理的基本步骤

困难的任务之一。复杂的分割过程导致成功解决要求物体被分别识别出来的成像问题需要大量处理工作。另一方面,不健壮且不稳定的分割算法几乎总是会导致最终失败。通常,分割越准确,识别越成功。

表示与描述几乎总是跟随在分割步骤的输出后边,通常这一输出是未加工的数据,其构成不是区域的边缘(区分一个图像区域和另一个区域的像素集)就是其区域本身的所有点。无论哪种情况,把数据转换成适合计算机处理的形式都是必要的。首先,必须确定数据是应该被表现为边界还是整个区域。当注意的焦点是外部形状特性(如拐角和曲线)时,则边界表示是合适的。当注意的焦点是内部特性(如纹理或骨骼形状)时,则区域表示是合适的。在某些应用中,这些表示方法是互补的。选择一种表现方式仅是解决把原始数据转换为适合计算机后续处理的形式的一部分。为了描述数据以使感兴趣的特征更明显,还必须确定一种方法。描述也叫特征选择,涉及提取特征,该特征是某些感兴趣的定量信息或是区分一组目标与其他目标的基础。

识别是基于目标的描述给目标赋以符号的过程。如1.1节详细讨论的那样,我们用识别个别目标方法的开发推出数字图像处理的覆盖范围。

到目前为止,还没有谈到图1.23中关于先验知识及知识库与处理模块之间的交互这部分内容。关于问题域的知识以知识库的形式被编码装入一个图像处理系统。这一知识可能如图像细节区域那样简单,在这里,感兴趣的信息被定位,这样,限制性的搜索就被引导到寻找的信息处。知识库也可能相当复杂,如材料检测问题中所有主要缺陷的相关列表或者图像数据库(该库包含变化检测应用相关区域的高分辨率卫星图像)。除了引导每一个处理模块的操作,知识库还要控制模块间的交互。这一特性在图1.23中的处理模块和知识库间用双头箭头表示。相反,单头箭头连接处理模块。

虽然没有专门地讨论图像显示,但图像显示是很重要的。图像处理结果的观察可在

图1.23中任何一级的输出执行。还注意到不是所有的图像处理应用都需要图1.23所指出的复杂交互。事实上,在某些情况下并不是所有那些模块都需要。例如,为了人的视觉解释,图像增强很少需要使用图1.23中的其他任何步骤。然而,通常随着图像处理任务复杂度的增加要求做更多处理才能解决问题。

1.5 图像处理系统的部件

20世纪80年代中叶,世界各地售出的各种型号的图像处理系统,确切地说,基本都是由主机及与主机相配合的外设装置构成。20世纪80年代末90年代初,市场已转为将图像处理硬件设计为与工业标准总线兼容并能配合工程工作站机箱和个人计算机的单板形式。除了降低价格外,这一市场还如催化剂一样出现了大量的新公司,这些公司的任务是开发用于图像处理的软件。

虽然针对大规模图像应用(如处理卫星图像)的大规模图像处理系统一直在出售,但趋势朝着小型化和通用化的小型机并带有专用图像处理软件的混合型系统方向发展。图1.24显示了用于数字图像处理的典型通用系统的组成部件。每一部分的功能在下面的段落从图像感知开始进行讨论。

关于感知,需要两个部件以获取数字图像。第一个是物理设备,该设备对我们希望成像的物体发射的能量很敏感。第二个称之为数字化器,数字化器是一种把物理感知装置的输出转换为数字形式的设备。例如,在数字视频摄像机中,传感器产生一个与光强成比例的输出,数字化器把该输出转化为数字数据。这一题目的某些细节包含在第2章中。

特殊的图像处理硬件通常由刚刚谈到的数字化器与执行其他原始操作[如算术逻辑单元(ALU)]的硬件组成,算术逻辑单元对图像执行并行操作。一个使用ALU的例子是与数字化一样快的图像取平均操作,这一操作的目的是降低噪声。有时这种类型的硬件叫做前端子系统,它的显著特点是速度快。换句话说,这个单元执行数据需要快速通过的功能[例如,在30帧/秒的速度下数字化和求视频图像的平均值],典型的主机不能胜任该工作。

在图像处理系统中的计算机是通用的计算机,其范围从PC机到超级计算机。有时在专门应用中,也采用特殊设计的计算机以达到所要求的性能水平。但是,这里感兴趣的还是通用图像处理系统。在这些系统中,几乎任何配置较好的PC机对于离线图像处理任务都是适合的。

图像处理软件由执行特殊任务的特殊模块组成。一个设计优良的软件包还包括为用户写代码的能力。如最小化就使用专用模块完成。完善的软件包允许那些模块和至少用一种计算机语言编写的通用软件命令集成。

大规模存储能力在图像处理中是必需的。一幅图像的尺寸是1 024×1 024像素,每像素的亮度是8比特,如果图像不压缩需1 MB的存储空间。当处理1千甚至1百万幅图像时,在图像处理系统中很难提供足够的存储空间。图像处理应用的存储分成三个主要类别:(1)用于处理时的短期存储;(2)关系到快速调用的在线存储;(3)档案存储,其特点是频繁访问。存储是以字节(B),千字节(KB),兆字节(MB),吉字节(GB),太字节(TB)为单位的。

提供短期存储的一种方法是使用计算机内存。另一种方法采用专门的板,该存储板叫做帧缓存(缓冲存储器),它们可以存储一帧或多帧图像并可快速访问,通常以视频速率

(30帧/秒)访问。后一种方法允许实际意义上的快速图像变焦以及卷动(垂直移动)和摇(水平移动)。帧缓存通常放在专业的图像处理硬件单元中,如图1.24所示。在线存储一般采取磁盘或光介质存储。在线存储的关键特性参数是对存储数据的访问频率。最后,档案存储是以海量存储要求为特点的,但无需频繁访问。放在类似于投币电唱机的盒子内的磁带和光盘是通常使用的档案存储介质。

今天使用的图像显示器主要是彩色电视监视器(更好一些的平面屏幕)。监视器由图像和图形显示卡驱动,它们是计算机系统的组成部分。对于图像显示应用不可能没有这样的要求,即作为计算机系统的一部分,其显示卡应满足商用性要求,在有些场合还要求立体显示。立体显示是采用戴在用户头上的头盔(目镜上嵌有两个小的显示屏)来实现的。

用于记录图像的硬拷贝装置包括激光打印机、胶片照相机、热敏装置、喷墨装置和数字单元(如CD-ROM)等。胶片的分辨率最高,但相纸作为记录材料是首要选择的介质。为了描述图像,图像可显示在透明胶片上或者使用图像投影设备显示在数字介质上。后面的方法作为图像描述标准正在被接受。

网络在今天所用的计算机系统中几乎都是默认的功能,因为大数据量在图像处理中是固有的,在图像传输中主要考虑的问题是带宽。在专用网络中这不是一个典型问题,但通过互联网远程通信就不是总有效的了。幸运的是,这一状况由于光纤和其他宽带技术的发展,正在迅速得到改进。

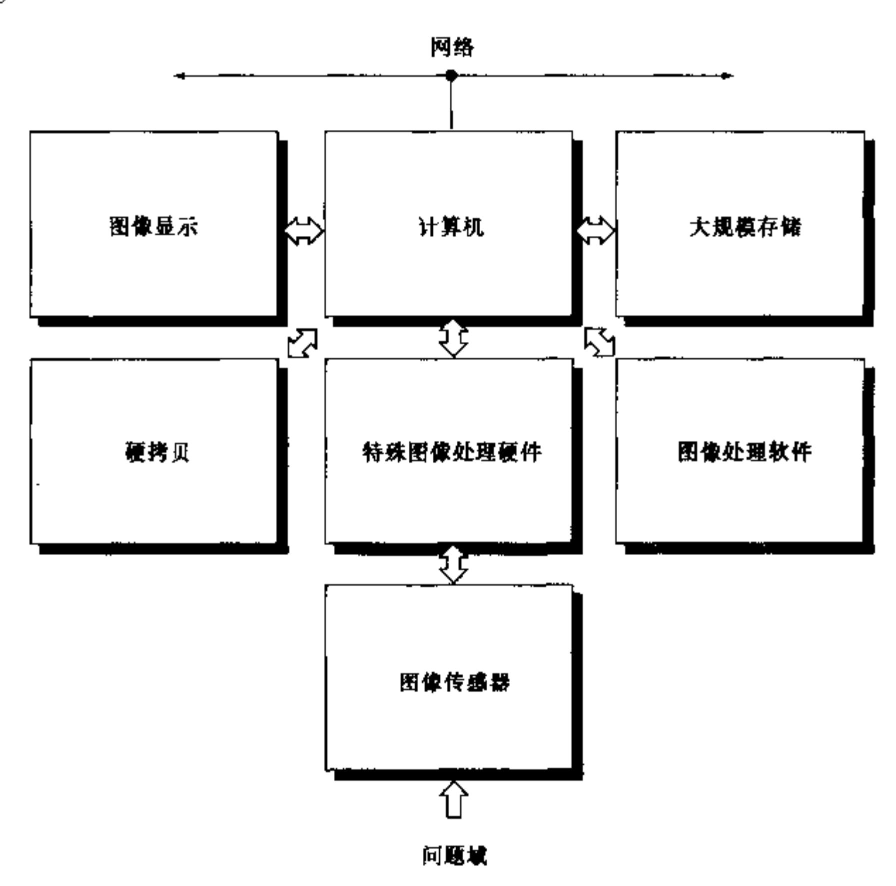


图1.24 通用图像处理系统组成

小结

本章提供的材料主要是,对数字图像处理的起源、重要意义、该技术当前及未来的应用领域提供一个概括性的介绍。由于篇幅所限,本章中涉及课题的覆盖面必然不太全面,但是在数字图像处理的宽度和应用范围方面应该给读者留下了一个总体印象。在后面的章节中,将进行图像处理的理论和应用方面的阐述,并提供大量的实例进一步理解这些技术的使用。学习完本书后,读者将对数字图像处理的基础内容有一个全面了解。

参考资料

位于后续各章末尾的参考资料部分列出了那些章节讨论过的文献,并在本书末尾的“参考文献”中统一列出。然而,为了在这一章集中汇总一下关于图像处理和相关题目的材料所出自的杂志,我们采取不同的形式。这里还提供了一个书目列表,以便读者可以快速了解本领域的历史和当前的发展动态。这一章引用的参考材料主要是图像处理方面的一般目的、容易理解的文献。

关于图像处理及相关题目的主要参考杂志包括: *IEEE Transactions on Image Processing*; *IEEE Transactions on Pattern Analysis and Machine Intelligence*; *Computer Vision, Graphics, and Image Processing* (1991 年前); *Computer Vision and Image Understanding*; *IEEE Transactions on Systems, Man and Cybernetics*; *Artificial Intelligence*; *pattern Recognition*; *Pattern Recognition Letters*; *Journal of the Optical Society of America* (1984 年前); *Journal of the Optical Society of America——A: Optics, Image Science and Vision*; *Optical Engineering*; *Applied Optics——Information Processing*; *IEEE Transaction on Medical Imaging*; *Journal of Electronic Imaging*; *IEEE Transactions on Information Theory*; *IEEE Transactions on Communications*; *IEEETra-nsactions on Acoustics, Speech and Signal Processing*; *Proceedings of the IEEE*; 以及 1980 年前的 *IEEE Transactions on Computer* 的文章。国际光学工程协会的出版刊物也很重要。

下面按年代反序列出的书籍包含了数字图像处理方面的补充材料(更偏重于近年来出版的书籍)。这些书籍描述了该领域在过去 30 年总的观点,同时选择了各种形式的论著。覆盖范围从基础理论的教科书到提供技术依据的手册,并包含了该领域当前研究的代表性材料与书籍。

Duda, R. O., Hart, P. E., and Stork, D. G. [2001]. *Pattern Classification*, 2nd ed., John Wiley & Sons, NY.

Pratt, W. K. [2001]. *Digital Image Processing*, 3rd ed., John Wiley & Sons, NY.

Ritter, G. X. and Wilson, J. N. [2001]. *Handbook of Computer Vision Algorithms in Image Algebra*, CRC Press, Boca Raton, FL.

Shapiro, L. G. and Stockman, G. C. [2001]. *Computer Vision*, Prentice Hall, Upper Saddle River, NJ.

Dougherty, E. R. (ed.) [2000]. *Random Processes for Image and Signal Processing*, IEEE Press, NY.

Etienne, E. K. and Nachtegael, M. (eds.). [2000]. *Fuzzy Techniques in Image Processing*,

Springer-Verlag, NY.

Goutsias, J, Vincent, L., and Bloomberg, D. S. (eds.). [2000]. *Mathematical Morphology and Its Applications to Image and Signal Processing*, Kluwer Academic Publishers, Boston, MA.

Mallot, A. H. [2000]. *Computational Vision*, The MIT Press, Cambridge, MA.

Marchand-Maillet, S. and Sharaiha, Y. M. [2000]. *Binary Digital Image Processing: A Discrete Approach*, Academic Press, NY.

Mitra, S. K. and Sicuranza, G. L. (eds.) [2000]. *Nonlinear Image Processing*, Academic Press, NY.

Edelman, S. [1999]. *Representation and Recognition in Vision*, The MIT Press, Cambridge, MA.

Lillesand, T. M. and Kiefer, R. W. [1999]. *Remote Sensing and Image Interpretation*, John Wiley & Sons, NY.

Mather, P. M. [1999]. *Computer Processing of Remotely Sensed Images: An Introduction*, John Wiley & Sons, NY.

Petrou, M. and Bosdogianni, P. [1999]. *Image Processing: The Fundamentals*, John Wiley & Sons, UK.

Russ, J. C. [1999]. *The Image Processing Handbook*, 3rd ed., CRC Press, Boca Raton, FL.

Smirnov, A. [1999]. *Processing of Multidimensional Signals*, Springer-Verlag, NY.

Sonka, M., Hlavac, V., and Boyle, R. [1999]. *Image Processing, Analysis, and Computer Vision*, PWS Publishing, NY.

Umbaugh, S. E. [1998]. *Computer Vision and Image Processing: A Practical Approach Using CVIPtools*, Prentice Hall, Upper Saddle River, NJ.

Haskell, B. G. and Netravali, A. N. [1997]. *Digital Pictures: Representation, Compression, and Standards*, Perseus Publishing, NY.

Jahne, B. [1997]. *Digital Image Processing: Concepts, Algorithms, and Scientific Applications*, Springer-Verlag, NY.

Castleman, K. R. [1996]. *Digital Image Processing*, 2nd ed., Prentice Hall, Upper Saddle River, NJ.

Geladi, P. and Grahn, H. [1996]. *Multivariate Image Analysis*, John Wiley & Sons, NY.

Bracewell, R. N. [1995]. *Two-Dimensional Imaging*, Prentice Hall, Upper Saddle River, NJ.

Sid-Ahmed, M. A. [1995]. *Image Processing: Theory, Algorithms, and Architectures*, McGraw-Hill, NY.

Jain, R., Rangachar, K., and Schunk, B. [1995]. *Computer Vision*, McGraw-Hill, NY.

Mitiche, A. [1994]. *Computational Analysis of Visual Motion*, Perseus Publishing, NY.

Baxes, G. A. [1994]. *Digital Image Processing: Principles and Applications*, John Wiley & Sons, NY.

Gonzalez, R. C. and Woods, R. E. [1992]. *Digital Image Processing*, Addison-Wesley, Reading, MA.

Haralick, R. M. and Shapiro, L. G. [1992]. *Computer and Robot Vision*, vols. 1 & 2, Addison-Wesley, Reading, MA.

Pratt, W. K. [1991] *Digital Image Processing*, 2nd ed., Wiley-Interscience, NY.

Lim, J. S. [1990]. *Two-Dimensional Signal and Image Processing*, Prentice Hall, Upper Saddle River, NJ.

Jain, A. K. [1989]. *Fundamentals of Digital Image Processing*, Prentice Hall, Upper Saddle River, NJ.

Schalkoff, R. J. [1989]. *Digital Image Processing and Computer Vision*, John Wiley & Sons, NY.

Giardina, C. R. and Dougherty, E. R. [1988]. *Morphological Methods in Image and Signal Processing*, Prentice Hall, Upper Saddle River, NJ.

Levine, M. D. [1985]. *Vision in Man and Machine*, McGraw-Hill, NY.

Serra, J. [1982]. *Image Analysis and Mathematical Morphology*, Academic Press, NY.

Ballard, D. H. and Brown, C. M. [1982]. *Computer Vision*, Prentice Hall, Upper Saddle River, NJ.

Fu, K. S. [1982]. *Syntactic Pattern Recognition and Applications*, Prentice Hall, Upper Saddle River, NJ.

Nevatia, R. [1982]. *Machine Perception*, Prentice Hall, Upper Saddle River, NJ.

Pavlidis, T. [1982]. *Algorithms for Graphics and Image Processing*, Computer Science Press, Rockville, MD.

Rosenfeld, R. and Kak, A. C. [1982]. *Digital Picture Processing*, 2nd ed., vols. 1 & 2, Academic Press, NY.

Hall, E. L. [1979]. *Computer Image Processing and Recognition*, Academic Press, NY.

Gonzalez, R. C. and Thomason, M. G. [1978]. *Syntactic Pattern Recognition: An Introduction*, Addison-Wesley, Reading, MA.

Andrews, H. C. and Hunt, B. R. [1977]. *Digital Image Restoration*, Prentice Hall, Upper Saddle River, NJ.

Pavlidis, T. [1977]. *Structural Pattern Recognition*, Springer-Verlag, NY, 1977.

Tou, J. T. and Gonzalez, R. C. [1974]. *Pattern Recognition Principles*, Addison-Wesley, Reading, MA, 1974.

Andrews, H. C. [1970]. *Computer Techniques in Image Processing*, Academic Press, NY.

第2章　数字图像基础

本章主要介绍一些与数字图像相关的概念和贯穿全书的一些标记符号。2.1节简单地综述了人类视觉系统的构造，包括眼中图像的形成及对亮度的适应和鉴别能力。2.2节讨论光、电磁波谱的其他分量和它们的成像特点。2.3节讨论成像传感器及怎样用它们产生数字图像。2.4节介绍均匀图像取样及灰度量化的概念。在这部分还讨论了数字图像表示、图像中取样数目和灰度级变化的影响、与取样有关的重要现象及图像放大、缩小的技术。2.5节处理一些在全书都要用到的像素间的基本关系。最后，2.6节定义了线性操作的条件。正如本节指出的，线性算子在图像处理技术中起着核心作用。

2.1　视觉感知要素

虽然数字图像处理领域建立在数学和概率统计表示法的基础上，但人的直觉和分析相比之下在选择一种技术时起到核心作用。这种选择常常是主观的视觉判断。因此，大概了解一下人类的视觉感知应作为学习全书的第一步。鉴于这一议题的复杂性和广度，我们仅仅涉及人类视觉最基本的方面。特别是，我们把兴趣放在人眼中形成图像的机理与参数上。这里将通过用于数字图像处理的一些因素去了解人类视觉的物理限制。例如，在分辨率和对照明变化的适应能力方面怎样对人类和电子图像进行比较，这些内容不仅有趣，而且从实践的观点看也是很重要的。

2.1.1　人眼的构造

图2.1显示了人眼简化了的横截面。眼睛的形状近似于一个圆球，其平均直径大约为20 mm。有三层薄膜包围着眼睛，即眼角膜和巩膜外壳，脉络膜和视网膜。角膜是一种硬而透明的组织，它覆盖着眼睛的前表面。与角膜相连的巩膜是一层包围着眼球剩余部分的不透明膜。

脉络膜位于巩膜的下面，这层膜包含有血管网，它是眼睛的重要滋养源。即使是对脉络膜表面的损害(这种损害常常认为并不严重)也有可能严重地损害眼睛。脉络膜外壳着色很重，因此，有助于减少进入眼内的外来光和眼球内反向散射光的数量。在脉络膜的最前面分为睫状体和虹膜。虹膜的收缩和扩张控制着进人眼睛的光量。虹膜中间开口处(瞳孔)的直径是可变的，大约在2～8 mm的范围。虹膜的前部有眼睛的可见色素，而后部则有黑色素。

晶状体是由同心的纤维细胞层组成并由附在睫状体上的纤维悬挂着。晶状体包含60%～70%的水，6%的脂肪和比眼睛中任何其他组织都多的蛋白质。晶状体由稍黄的色素着色，其颜色随着年龄的增加而加深。在极端情况，晶状体会过于混浊，这通常是由白内障等疾病引起的，可能导致低下的彩色辨别能力和视觉清晰度的损失。晶状体吸收大约8%的可见光谱，对短波长光有较高的吸收率。在晶状体结构中，蛋白质吸收红外光和紫外光，并且过量会伤害眼睛。

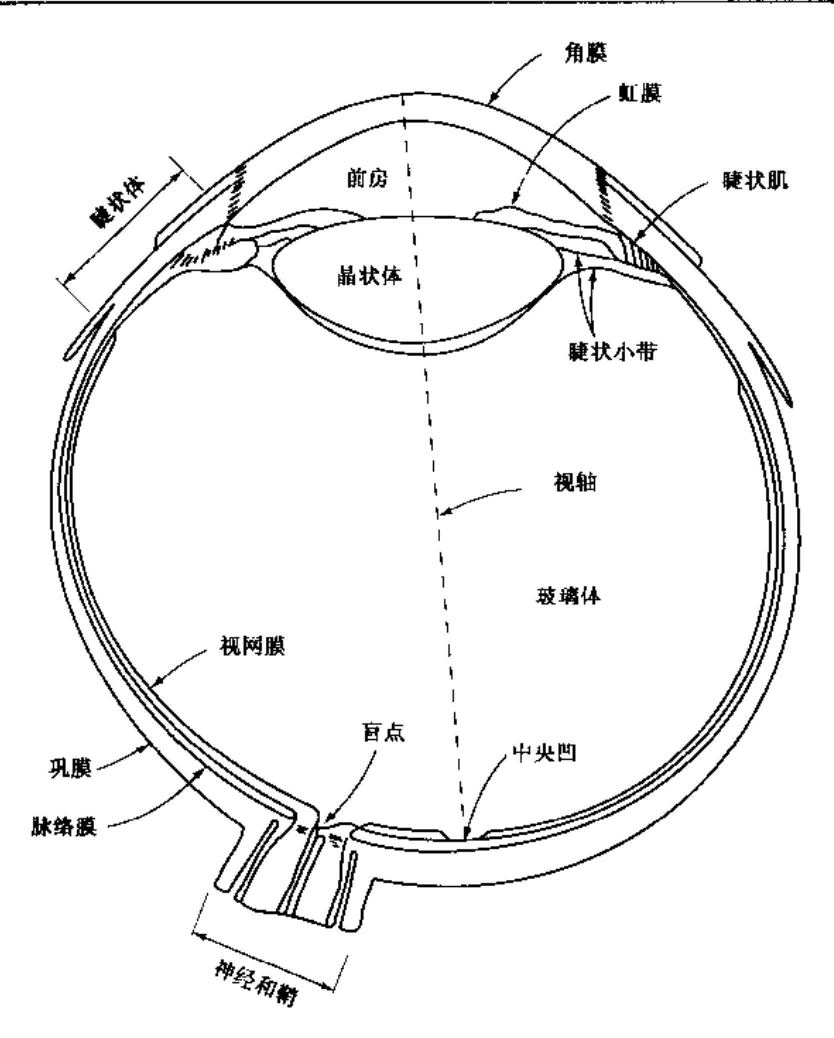


图 2.1　人眼横截面简图

眼睛最里面的膜是视网膜,它布满了整个后部的内壁。当眼球适当地聚焦时,来自眼睛外部的光在视网膜上成像。在视网膜表面分布的分离光接收器提供了图案视觉。这种光接收器分为两类:锥状体和杆状体。每只眼睛的锥状体数目在 600 万到 700 万之间。它们主要位于视网膜的中间部分,称之为中央凹,且对颜色灵敏度很高。用这些锥状体,人们可以充分地分辨图像细节,因为每一个锥状体都连接到自身的神经末端。肌肉控制眼球转动,直到感兴趣的物体图像落到中央凹上。锥状视觉叫做白昼视觉或亮光视觉。

杆状体数目更多,约有 7 500 万到 15 000 万个杆状体分布在视网膜表面。由于分布面积较大而且几个杆状体连接到一个神经末端,因此,减少了这些接收器感知细节的数量。杆状体用来给出视野内一般的总体图像。它们没有彩色感觉,而在低照明度下对图像较敏感。例如,在白天呈现鲜明色彩的物体,在月光下都没有颜色,因为此时只有杆状体受到刺激。这个现象就是众所周知的夜视觉或叫做暗视觉。

图 2.2 显示了右眼中通过眼睛光神经应急区的横截面的杆状体和锥状体密度。在这一区域没有接收器导致所谓的盲点(见图 2.1)。除了这一区域,接收器的分布是沿视线关于中央凹对称的。接收器密度从中央凹的度数来度量(即:离开轴的度数由视轴和通过晶状体中心相交于视网膜的直线形成的角度决定)。注意,图 2.2 中锥状体在视网膜中心最密(在中央凹的

中心区)。应注意到,杆状体从中心向外,离轴大约 20°处,密度逐渐增加,然后向外到视网膜的边缘处密度逐渐下降。

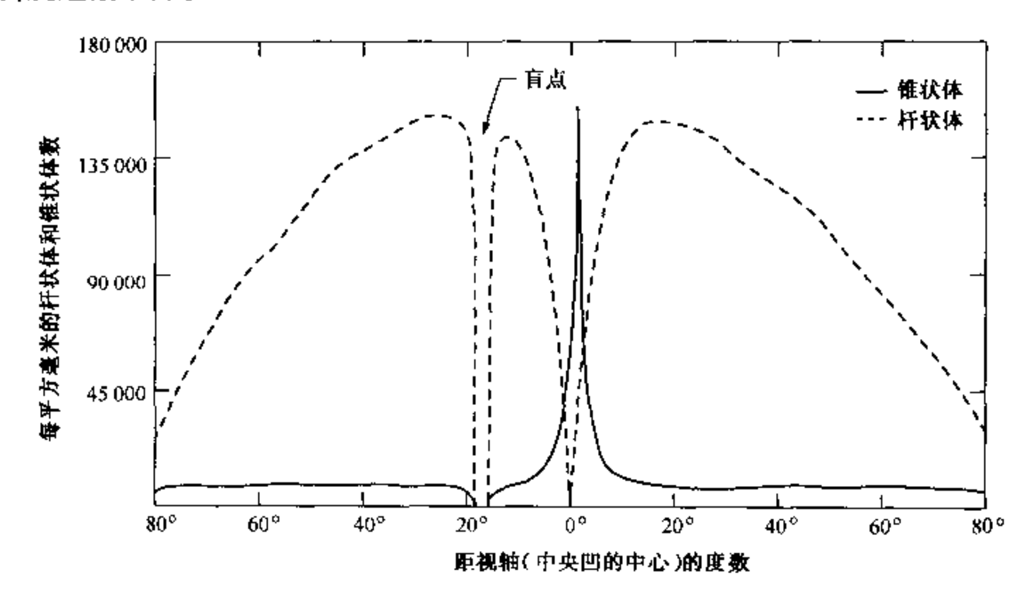


图 2.2　视网膜上杆状体和锥状体的分布

中央凹本身是在视网膜上直径约 1.5 mm 的凹坑。然而,在后续的讨论中,把它近似为方形或矩形的敏感元素的阵列更有用。这样,在描述上比较灵活,可以把中央凹作为大小为 1.5 mm×1.5 mm的方形传感器阵列观察。在视网膜这一区域中,锥状体的密度大约为每平方毫米 150 000 个元素。基于这一近似,眼睛中最高敏感区域的锥状体数量约为 337 000 个元素。从行分辨能力的观点看,恰好是一个电荷耦合元件(CCD)中等分辨率的成像芯片可有这个数量的元素,且接收器阵列不大于 5 mm×5 mm。尽管人类的智慧和视觉经验使这种比较并不恰当,但在当前电子成像传感器领域进一步讨论人眼分辨细节的基本能力时,这种比较是应肯定的。

2.1.2　眼睛中图像的形成

眼睛的晶状体和普通光学透镜之间的主要差别在于前者的适应性强。正如图 2.1 所表明的,晶状体前表面的曲率半径大于后表面的曲率半径。晶体状的形状由睫状体韧带和张力来控制。为了对远方的物体聚焦,控制肌肉使晶状体相对比较扁平。同样,为对眼睛近处的物体聚焦,肌肉会使晶状体变得较厚。

当晶状体的折射能力由最小变到最大时,晶状体的聚焦中心与视网膜间的距离由 17 mm 缩小到 14 mm。当眼睛聚焦到远于 3 m 的物体时,晶状体的折射能力最弱。当眼睛聚焦到非常近的物体时,晶状体的折射能力最强。这一信息使得很容易计算出任何图像在视网膜上形成图像的大小。例如,图 2.3 中,观察者正在看一棵高 15 m,距离 100 m 的树。如果 h 为物体在视网膜上图像的高,单位为毫米,由图 2.3 的几何形状可以看出 $15/100 = h/17$ 或者 $h = 2.55$ mm。正像 2.1.1 节指出的那样,视网膜图像主要反射在中央凹区域上。然后,由光接收器的相应刺激作用产生感觉。感觉把辐射能转变为电脉冲,最后由大脑解码。

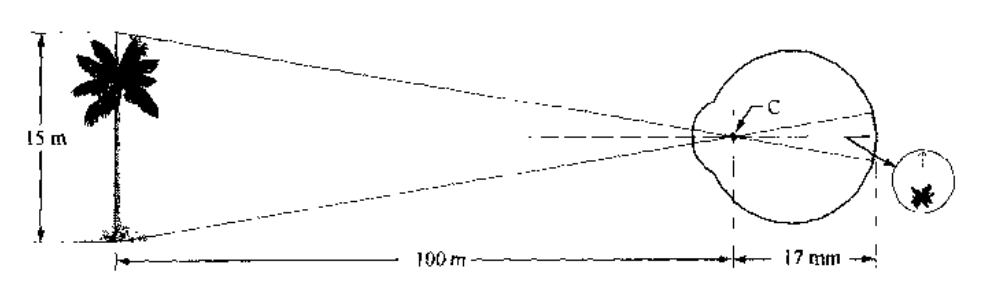


图 2.3 用眼睛看一棵棕榈树的图解,C 点是晶状体的光心

2.1.3 亮度适应和鉴别

因为数字图像作为离散的亮点集显示,所以,眼睛对不同亮度之间的鉴别能力在图像处理结果中是要考虑的重要方面。人的视觉系统能够适应的光强度级别范围是很宽的。从夜视阈值到强闪光约有 10^{10} 量级。实验数据指出,主观亮度(即由人的视觉系统感觉到的亮度)是进入眼睛的光强度的对数函数。图 2.4 中用光强度与主观亮度的关系曲线说明了这一特性。长的实线代表人的视觉系统能适应的光强度范围。昼视觉的范围是 10^6。由夜视觉到昼视觉逐渐过渡,过渡范围大约为从 0.001 到 0.1 mL(在对数标度中为 -3 到 -1 mL),图中画出了这一适应曲线的范围。

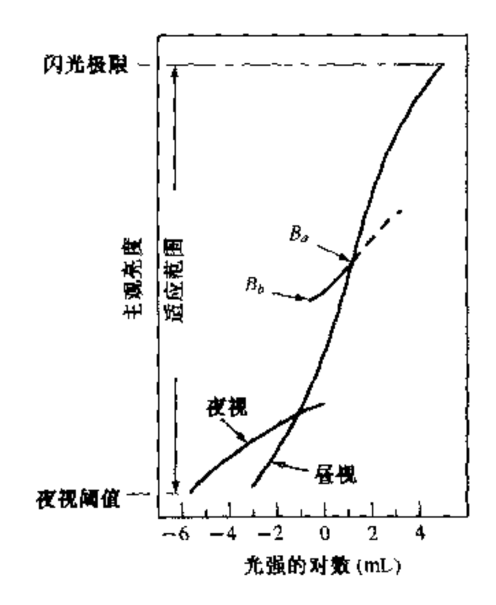


图 2.4 特殊适应水平的主观亮度感觉范围

解释图 2.4 特殊动态范围的基本点是人的视觉绝对不能同时在一个范围内工作,确切地说,它是利用改变其整个灵敏度来完成这一大变动的,这就是所谓的亮度适应现象。与整个适应范围相比,能同时鉴别的光强度级的总范围很小。对于任何一组给定条件,视觉系统当前的灵敏度级别叫做亮度适应级,例如,它可能相当于图 2.4 中的亮度 B_a。短的交叉线表示当眼睛适应这一强度级时人眼能感觉的主观亮度范围。注意,这一范围是有一定限制的,在 B_b 处和 B_b 以下时,所有的刺激都是作为不可分辨的黑色来理解。曲线的上部(虚线)实际上没有限

制，如果延伸太远，就会失去意义，因为高得多的强度将会把适应能力提高到比 B_a 更高的数值。

在任何特定的适应级，人眼辨别光强度之间变化的能力也是值得考虑的。用以确定人类视觉系统亮度辨别能力的一个典型实验由一个注视对象和均匀的、大到足以使其占有全部视野的发光区组成。这一区域是典型的漫反射体，例如，不透明玻璃，它被一个强度 I 可变的后光源照射时发光。在这一区域加上一个照射分量 ΔI，形成一个短期闪烁，该闪烁以均匀光场中央的圆形方式出现，如图 2.5 所示。

如果 ΔI 不够亮，则目标不变，表明没有可察觉的变化。当 ΔI 逐渐加强时，则目标给出一个正的响应，指出一个觉察到的变化。最后，ΔI 足够强，物体将始终给出"肯定"的响应。$\Delta I_c/I$ 称为韦伯比，这里 ΔI_c 是在背景照明为 I 的情况下可辨别照明增量的 50%。$\Delta I_c/I$ 值较小意味着可辨别强度较小的百分比变化，这表示亮度辨别能力好。反之，$\Delta I_c/I$ 较大意味着要求有较大百分比的强度变化，这表示亮度辨别能力较差。

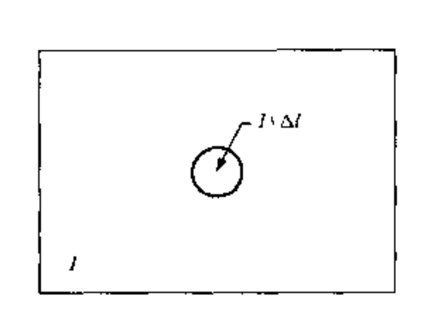


图 2.5　用于描述亮度辨别特性的基本实验

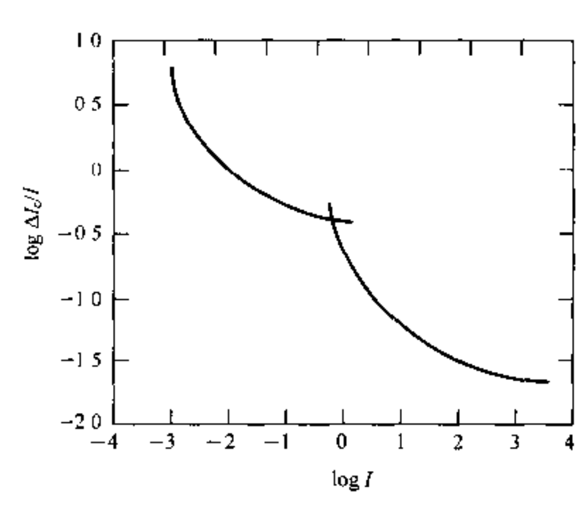


图 2.6　作为强度函数的典型韦伯比

作为 I 的对数函数，$\log \Delta I_c/I$ 曲线有图 2.6 所示的通常形状。这一曲线表明，在低的照明级别，亮度辨别较差（韦伯比大），当背景照明增加时亮度辨别得到有意义的改善（韦伯比降低）。曲线中的两个分支反映了这样的事实，即，在低照明级别情况，视觉由杆状体执行，在高照明级别情况，视觉由锥状体起作用（表示较好的辨别能力）。

如果背景照明保持恒定，并且代替闪光的其他光源的亮度从不能觉察到总可以被觉察间逐渐变化，一般观察者可以辨别总共 12 到 24 级不同强度变化。粗略地看，这个结果与一个人观看一幅单色图像的任一点时能觉察到的不同强度数量有关。这个结果并不意味着一幅图像可以用这样小的强度数值来表现，因为，当眼睛扫视图像时，平均背景在变化，这样，允许在每个新的适应水平上检测不同的增量变化。最后结果是眼睛能够辨别很宽的全部强度范围。事实上，正如 2.4.3 节所示，眼睛可以检测到用不足 24 级的全部强度描述的单色图像中很糟的轮廓效应。

两个现象可以证明感觉亮度不是简单的强度函数。第一个现象基于视觉系统倾向不同强度区域边界周围的"欠调"或"过调"。图 2.7(a)显示了这种现象的一个典型例子。虽然条带强度恒定，但实际感觉到了一幅带有毛边（特别是靠近边界处）的亮度图形[图 2.7(b)]。这些

表面上的毛边带叫做马赫带,是厄恩斯特·马赫在 1865 年首先描述的现象。

第二个现象称为同时对比现象,这与下面的事实有关。即,感觉的亮度区域不是简单地取决于强度,如图 2.8 所示,所有的中心方块都有完全相同的强度。然而,它们出现在眼睛里就变得比较暗,如同背景亮了一样。一个更熟悉的例子是一张纸,当放在桌子上时看上去似乎比较白,但是,当我们用纸来遮蔽眼睛直视明亮的天空时,纸看起来总是黑的。

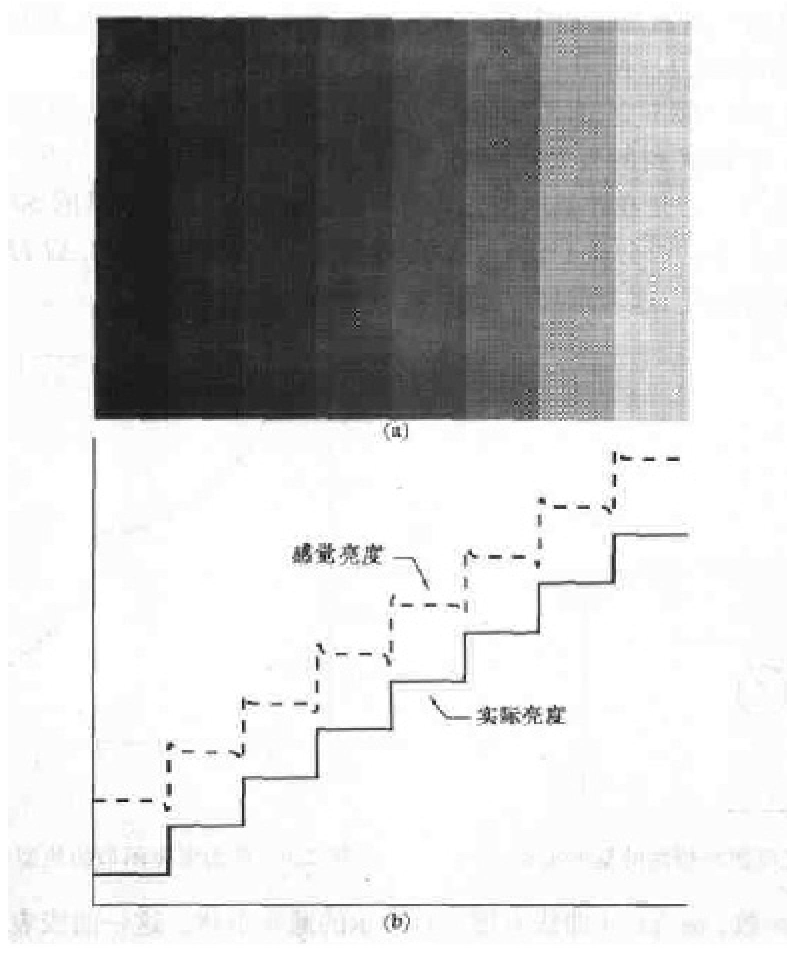


图 2.7 (a)显示感觉亮度不是简单的强度函数的一个例子,在(b)中两条曲线间相对垂直的位置没有特殊含意,只是为了清晰

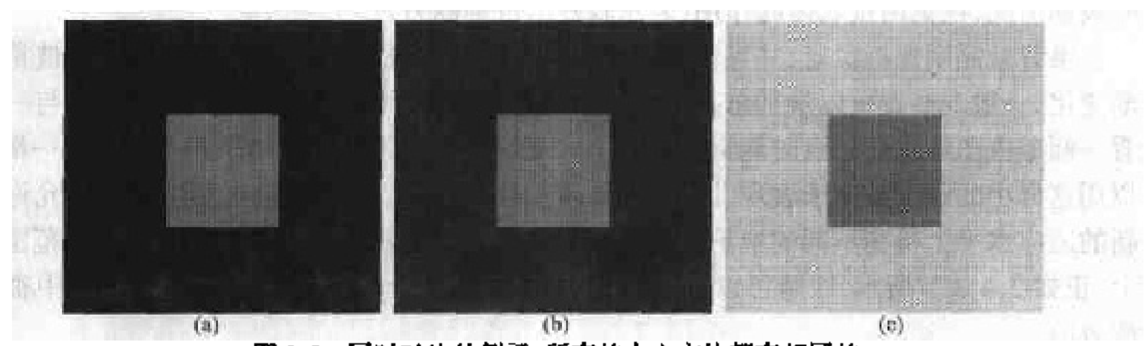


图 2.8 同时对比的例子,所有的中心方块都有相同的强度,但是当背景变亮时,它们就逐渐变暗

人类感知现象的另外一个例子是视觉错觉。在错觉中,眼睛填充上了不存在的信息或者错误地感知物体的几何特点。如图2.9(a)所示,一个正方形的轮廓看得很清楚,尽管在图像中这些部分并没有线。相同的效果在图2.9(b)中也可以看到,此时可以看到一个圆,注意到,仅仅是几根线就足以给出一个完整圆的错觉。在图2.9(c)中两条水平线段长度相同,但是一个显得比另一个短。最后,图2.9(d)中所有旋转45°的线都是等距离的平行线,然而画有交叉影线就产生了错觉,觉得这些线不再平行。视觉错觉是人类视觉系统的一个特性,这一特性尚未完全了解。

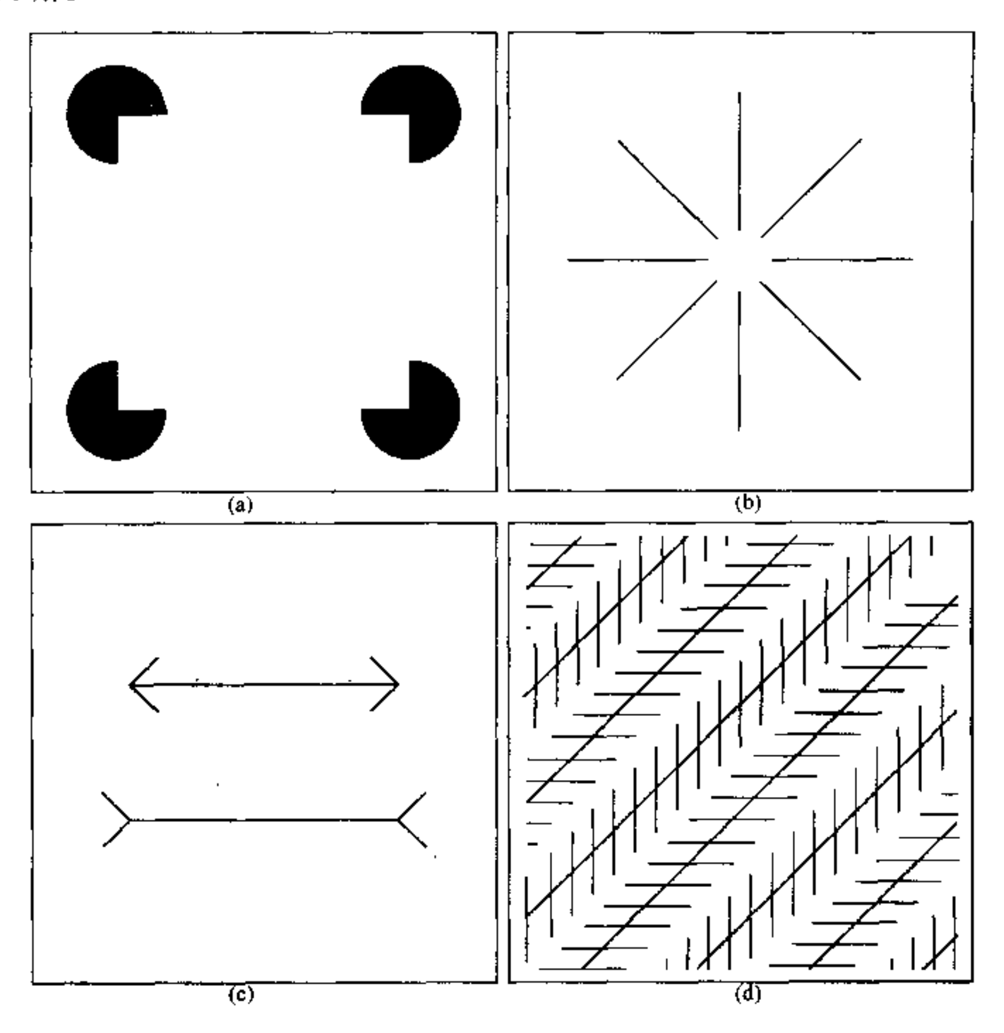


图2.9 一些典型的视觉错觉

2.2 光和电磁波谱

电磁波谱在1.3节中已做过介绍,现在要详细地研究这一问题。1666年,牛顿发现了这样一个现象:当一束太阳光通过一个玻璃棱镜时,显示的光束不再是白光,而是一端是紫色另一端是红色的连续色谱。如图2.10所示,我们感受到的可见光的彩色范围占电磁波的一小部分。在波谱的一端是无线电波,其波长是可见光波长的几十亿倍。波谱的另一端是伽马射线,其波长比可见光小几百万倍。电磁波谱可用波长、频率或能量来描述。波长(λ)和频率(ν)的

关系可用下式描述：

$$\lambda = \frac{c}{\nu} \tag{2.2.1}$$

这里 c 是光速(2.998×10^8 m/s)，电磁波谱的各种分量由下式给出：

$$E = h\nu \tag{2.2.2}$$

这里 h 是普朗克常数。

波长的单位是米(m)，也常用微米(μm，1 μm = 10^{-6}m)和纳米(nm，1 nm = 10^{-9}m)来表示。频率用赫兹(Hz)来度量，1 赫兹表示正弦波每秒 1 个周期。常用的能量单位是电子伏特(eV)。

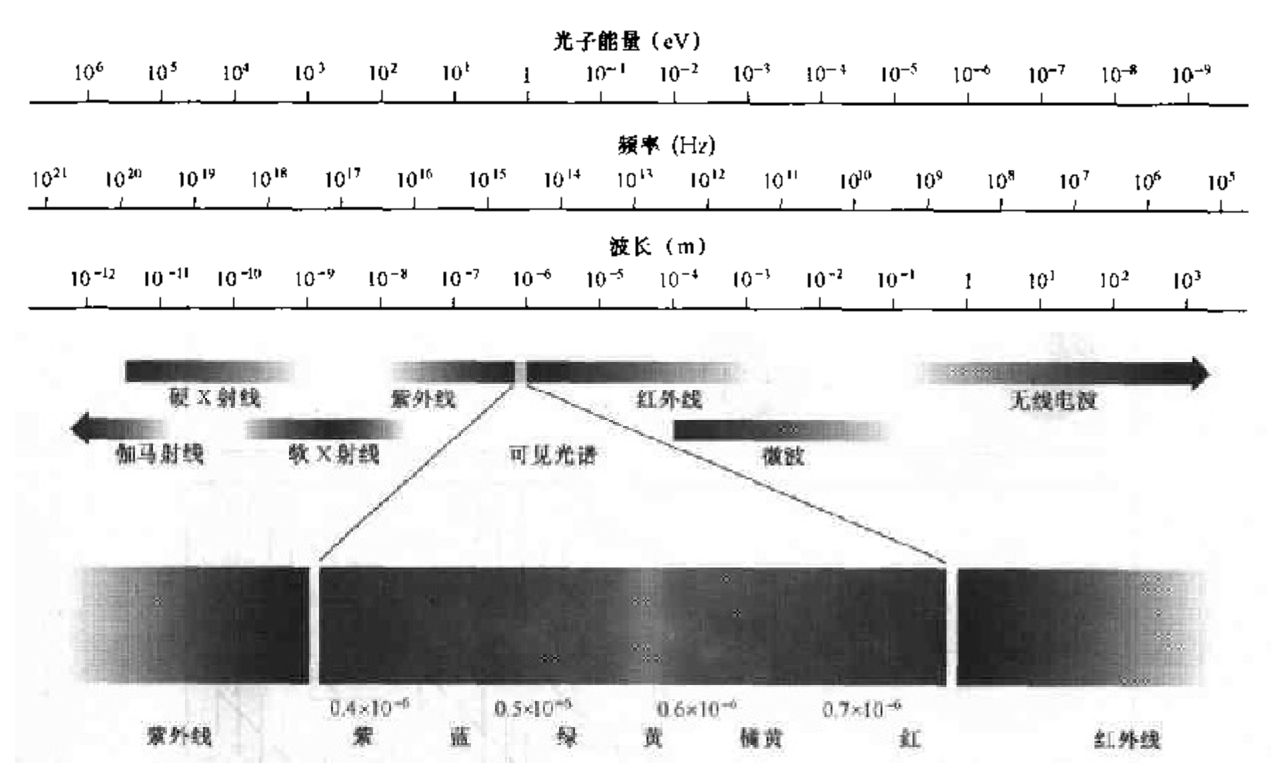


图 2.10 光和电磁波谱图

电磁波可看成是以波长 λ(图 2.11)传播的正弦波，或者看做没有质量的粒子流，该流以波那样的光速传播，每个粒子包含一定(一束)能量，每束能量称为光子。从式(2.2.2)可以看出能量与频率成正比，因此，高频率(短波长)电磁波每光子携带更多的能量，无线电波有低能量光子，微波具有比无线电波多的能量，红外波还要多。然后，可见光、紫外线、X 射线和最后的伽马射线都具有最高的能量。这就是伽马射线对活体组织危害很大的原因。

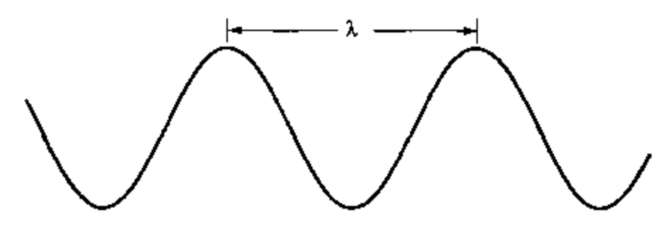


图 2.11 一个波长的图形表示

光是一种特殊的电磁辐射，人眼可以看见和感觉到它。为便于讨论，把可见光谱(彩色)段

扩展后示于图 2.10(在第 6 章将更详细地讨论彩色)中。电磁波谱的可见波段大约占据 0.43 μm(紫色)~0.79 μm(红色)的范围。为方便起见,彩色谱分为 6 个波段:紫色、蓝色、绿色、黄色、橘黄色和红色。每个波段(或电磁波谱的其他部分)不是突然终止,而是每一个波段混合平滑地过渡到下一个,如图 2.10 所示。

人从一个物体感受的颜色由物体反射光的性质决定。一个物体若所有反射的可见光波长是相对平衡的,则物体对观察者来说是显示白色。然而,一个物体在可见光谱的有限范围内反射时会呈现各种颜色色调。例如,绿色物体反射波长主要在 500~570 nm 范围的光,而吸收其他波长的大部分能量。

没有颜色的光叫消色或单色光。这种光的属性是它的强度或大小。灰度级这一词汇通常用来描述单色光强度,因为它的范围从黑到灰,最后到白。正如前面指出的,彩色光跨跃 0.43 μm(紫色)~0.79 μm(红色)的电磁能谱。通常有三个基本量用于描述彩色光源的质量:发光强度、光通量和亮度。发光强度是从光源流出能量的总量,通常用瓦特(W)度量。用流明数(lm)度量的光通量给出观察者从光源感受到的能量。例如,从远红外光谱范围的光源发射出来的光具有实际意义的能量,但观察者却很难感受到它。它的光通量几乎是零。最后,如 2.1 节讨论的那样,亮度是光感受的主观描绘子,它实际上不能测量,包含无色的强度概念,并且也是描述彩色感觉的参数之一。

接着从图 2.10 的讨论,我们注意到,在电磁波谱短波长的一端是伽马射线和硬 X 射线。如 1.3.1 节讨论的那样,伽马辐射对医学和天文学图像很重要,并且对核环境中辐射成像也很重要。硬(高能)X 射线用于工业。胸透 X 射线在软 X 射线区的高能端(短波长),牙科 X 射线在该波段的低能端。软 X 射线波段过渡到远紫外区,又与较长波长的可见光谱混合。进一步向较高波长移动就到达能辐射热的远红外波段,这一现象使它在依靠"热特性"成像的应用中非常有用。红外波段靠近可见光谱的部分叫近红外区。这个波段的对端叫做远红外区。远红外的后部与微波段相接。这个波段是众所周知的微波炉中的能源,但它有许多其他用途,包括通信和雷达。最后,无线电波段包括电视、AM 和 FM 收音机。从一定天体发射的高能无线电信号在天文观测中很有用。1.3 节给出了刚才讨论的多数波段成像的例子。

在原理上,如果可以开发出一种传感器,它可检测由一种电磁波谱发射的能量,就可以在那一波段上对感兴趣的事件成像。但有一点很重要,即"观看"一个物体的电磁波的波长必须小于或等于物体的尺寸。例如,水分子的直径是 10^{-10} m,这样,要研究分子,我们需要一个能在远红外或软 X 射线范围发射的光源。这个限制与传感材料的物理特性一起确立了成像传感器(如可见光、红外线和今天使用的其他传感器)功能的基本限制。

虽然成像主要以电磁波发射的能量为基础,但这并不是产生图像的惟一方法,例如,正如 1.3.7 节讨论过的从物体反射的声波可用于形成超声波图像。另外,主要的数字图像源是电子显微镜的电子束和用于图形和可视化的合成图像。

2.3　图像感知和获取

我们感兴趣的各类图像都是由"照射"源和形成图像的"场景"元素对光能的反射或吸收相结合而产生的。把"照射"和"场景"加上引号为了强调是比我们所熟悉的情况(即,可见光源照射各种三维场景)更一般的事实。例如,照射可能由电磁能源引起,如雷达、红外线或 X 射线

能源。但是,正如前面指出的,照射也可以由非传统光源,如超声波,甚至由计算机产生的照射模式产生。类似地,场景可能是熟悉的物体,但是,它们也可能是分子、沉积岩或人类大脑,甚至可以对一个光源成像,如获得太阳的图像。根据光源的性质,照射能量可从物体反射或从物体透射。第一类例子是从平面反射,第二类例子是使 X 光通过病人身体产生一幅 X 光诊断图像。在某些应用中,反射和透射都能聚焦到一个照片转换器上(图像屏幕),该转换器把能量转换为可见光。电子显微镜和伽马成像的应用就使用这种方法。

图 2.12 显示了用来把照射量变为数字图像的三种主要传感器装置。原理很简单,通过将输入电功率和对特殊类型检测能源敏感的传感器材料组合,把输入能源转变为电压。输出电压波形是传感器的响应,同时,一个数字量可从数字化该响应的每个传感器得到。在这一节,将看到图像感知和产生的主要程序,图像数字化将在 2.4 节讨论。

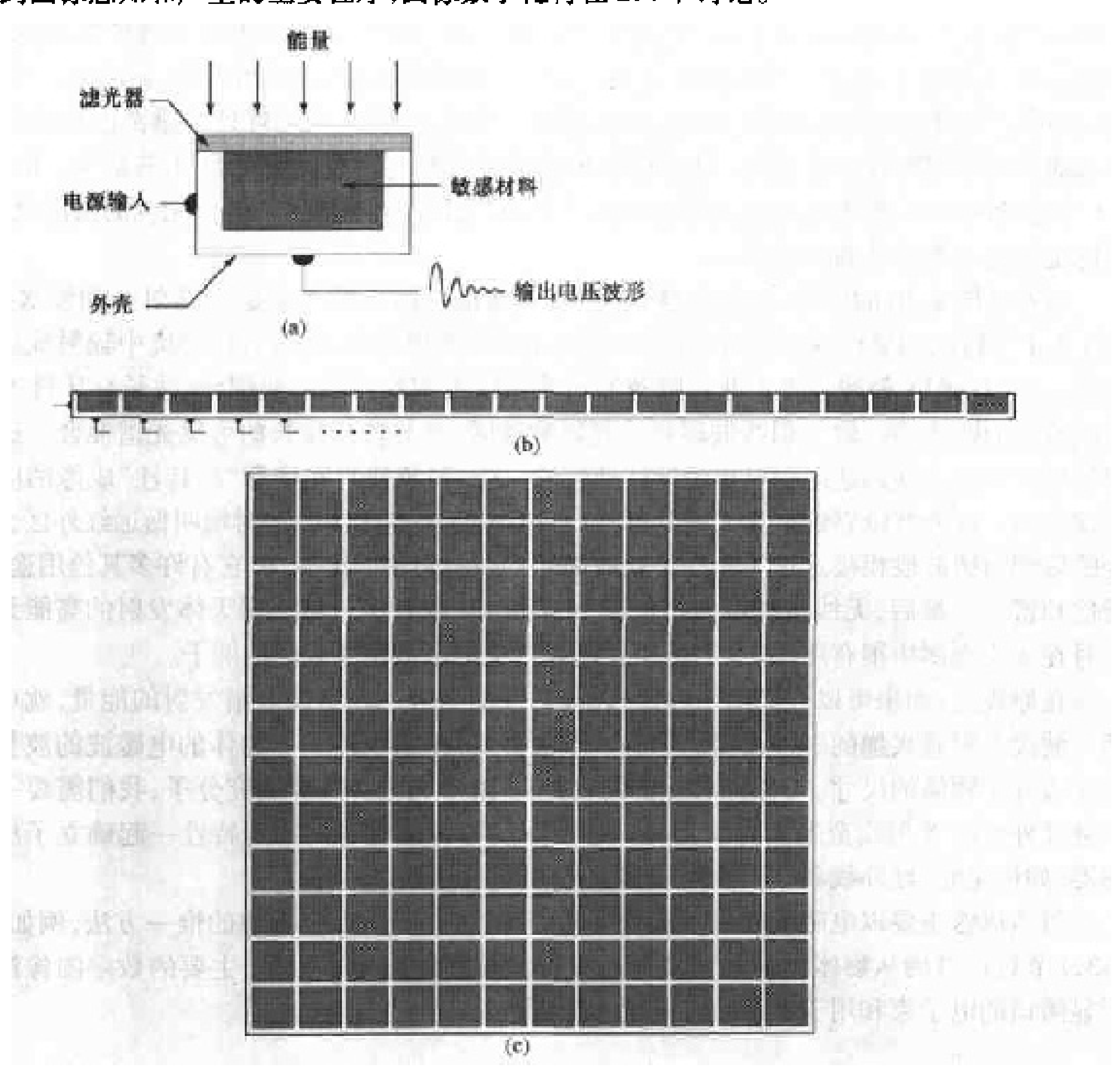


图 2.12　(a)单个成像传感器,(b)线性传感器,(c)传感器阵列

2.3.1　用单个传感器获取图像

图 2.12(a)显示了单个传感器部件。也许最熟悉的这类传感器是光二极管,它由硅材料构成,并且其输出电压波形与光成正比。在传感器前边用一个滤光器改善选择性。例如,在光传

感器前边的绿色(通过)滤光器有利于彩色谱的绿波段光通过。因此,传感器输出的绿光比可见光谱中其他分量要强。

为了用单一传感器产生二维图像,在传感器和成像区之间,还必须在 x 和 y 方向有相对位移。图 2.13 显示了一个用于高精度扫描的装置,这里把一个负片装在一只鼓上,该鼓机械转动提供一维位移。单传感器安装在引导螺杆上,它提供垂直方向上的移动。因为机械运动可高精度地控制,这一方法是得到高分辨率图像的一种廉价方法(但较慢)。另外,类似的机械装置使用一个平面床,而传感器在两个方向线性地移动,这些类型的机械数字化器有时归为微观光密度计一类。

用单一传感器成像的另外一个例子是同时放一个激光源和传感器,用一个镜子来控制发射光束到扫描图形上,同时把激光信号反射到传感器。这个装置也可使用条形或矩形传感器获取图像。这种装置将在后续两节加以讨论。

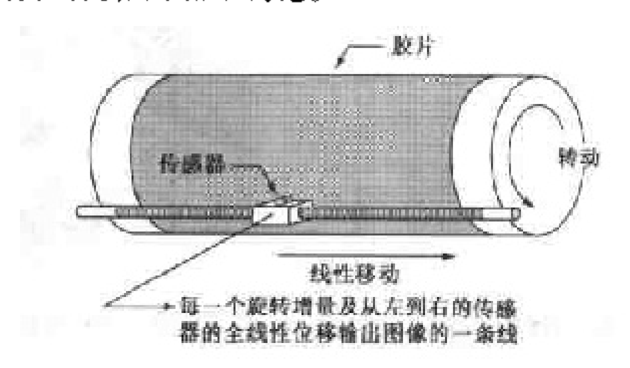


图 2.13　单个传感器组通过运动产生二维图像

2.3.2　用带状传感器获取图像

比单个传感器更常用的几何结构由线状排列的传感器形成的一个传感器带组成,如图 2.12(b)所示,传感器带在一个方向上提供成像单元。相对于传感器带垂直方向的运动在另一方向上成像,如图 2.14(a)所示,这是大多数平板扫描仪所用的装置。感受装置可能有 4 000或更多的排成直线的传感器。线状传感器常用于航空成像,在这种应用中,成像系统装在一架飞机上,飞机以恒定的高度和速度飞过被成像的地区。可响应各种电磁波谱波段的一维传感器带按与飞机方向垂直的方向安装。成像传感器带一次给出一幅图像的一行,传感器带的运动完成二维图像的另一维。透镜和其他聚焦方法用来把扫描区域投影到传感器上。

以圆环形状安装的传感器带用于医学和工业成像,以得到三维物体的横断面(切片)图像,如图 2.14(b)所示。一个旋转的 X 射线源提供照射,并且相对于射线源的传感器部分收集通过物体的 X 射线能量(很明显,传感器必须是对 X 射线敏感的)。这就是如 1.2 节和 1.3.2 节指出的医学和工业计算机轴向断层成像的基础。重要的一点是,传感器的输出必须由重建算法处理,重建的目的是把感知数据转换为有意义的横截面图像。换句话说,图像不可能从传感器的单一运动直接得到,它们需要进一步处理。由图像堆积组成的数字体由物体与传感器环相垂直方向的运动产生。基于 CAT 原理,其他成像模式包括核磁共振成像(MRI)和正电子发射断层(PET)。照射源、传感器和各种类型的图像是不同的,但是,在概念上,它们的成像方法[示于图 2.14(b)]是非常相似的。

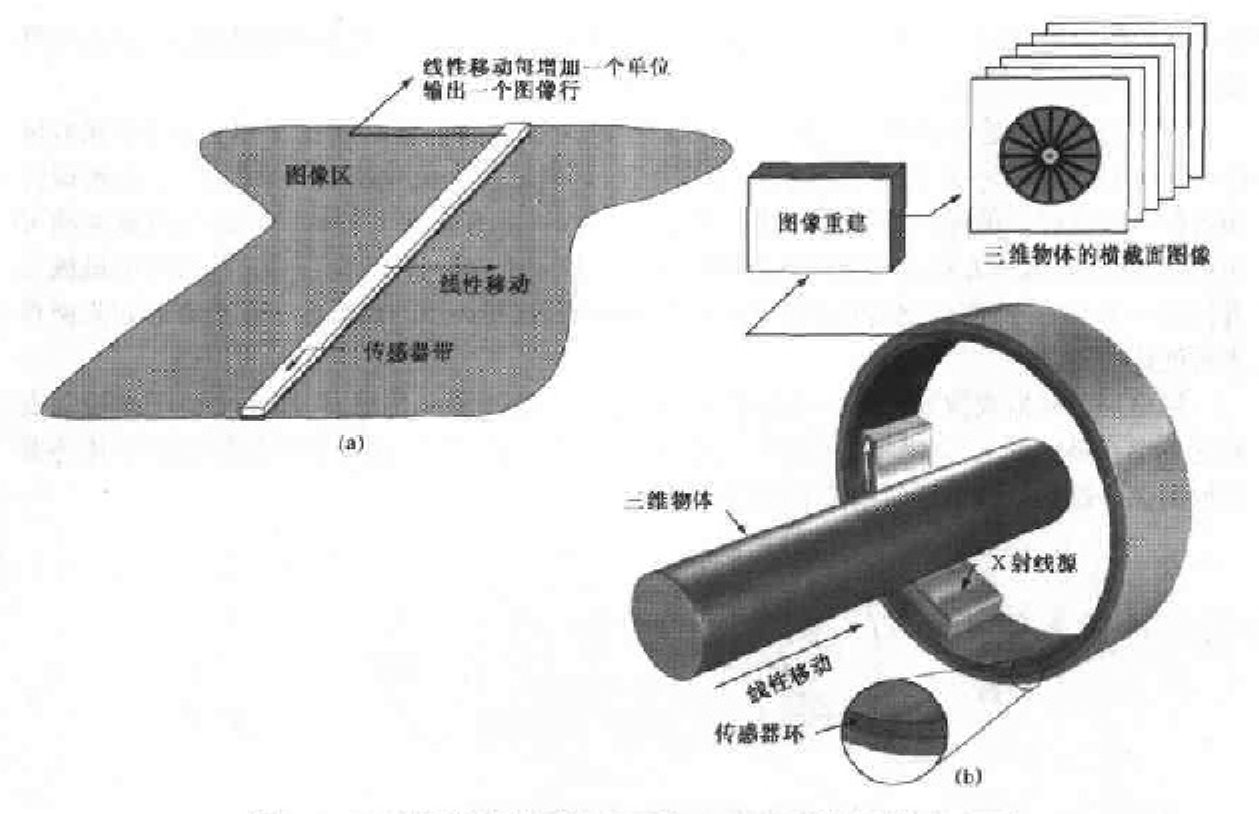


图 2.14 (a)用线性传感器带获取图像,(b)用环形传感器带获取图像

2.3.3 用传感器阵列获取图像

图 2.12(c)显示了独特的以二维阵列形式排列的传感器。大量的电磁波和某种超声波敏感元件常以阵列形式排列。这也是在摄像机上常见的主要结构。这些摄像机的典型传感器是CCD阵列。该种阵列可用宽带且具有敏感特性的元件制造并封装为4 000×4 000或更多单元的稳定阵列。CCD传感器广泛地运用于数字摄像机和其他光敏感设备中。每一个传感器的响应正比于投射到传感器表面的光能总量。这一特点被用于天文学和其他要求低噪声图像的应用中。采用让传感器累积输入的光信号超过几分钟甚至几小时的方法可达到减少噪声的目的(在第3章讨论用累积法减少噪声)。图 2.15(c)所示的传感器阵列是二维的,它的主要优点是,把图形能量聚焦到阵列表面得到完整的图像。很明显,像前面两节讨论的传感器结构那样的运动是不需要的。

在这样的传感器阵列中应用的主要方法示于图 2.15。该图显示了来自照射源通过场景(单元)反射的能量,但是,正如本节开始提到的,该能量也可以通过场景单元传送。示于图 2.15(c)的成像系统的第一种功能是收集入射能量并把它聚焦到一个图像平面上。如果照射的是光,则成像系统终端前面是一个透镜,透镜把观察场景投影到透镜的聚焦平面上,如图 2.15(d)所示。与焦点面相重合的传感器阵列产生与每一个传感器接收光总量成正比的输出。数字成模拟电路扫描这些输出,并把它们转换成电视信号,然后由成像系统的其他部分数字化,如图 2.15(e)所示输出数字图像。把一幅图像转换为数字图像是 2.4 节的话题。

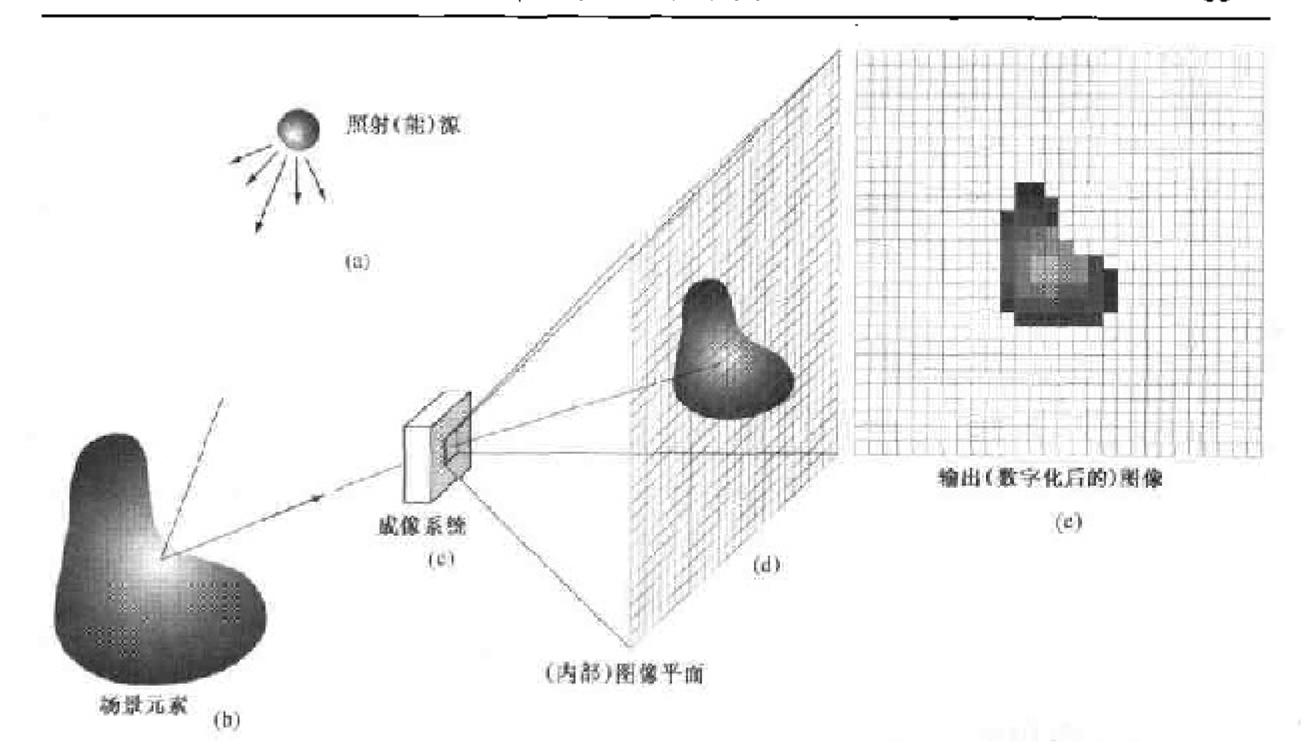


图 2.15　数字图像获取过程的例子。(a) 照射(能)源,(b)场景元素,(c)成像系统,(d)场景投影到图像平面,(e)数字化图像

2.3.4　简单的图像形成模型

用 $f(x,y)$二维函数形式表示图像,在特定的坐标(x,y)处,f 的值或幅度是一个正的标量,其物理意义由图像源决定。在本书中的大多数图像是单色图像,图像值如 2.2 节讨论的那样可称为灰度级。当一幅图像从物理过程产生时,它的值正比于物理源的辐射能量(如电磁波)。因此,$f(x,y)$一定是非零和有限的,这就是:

$$0 < f(x,y) < \infty \tag{2.3.1}$$

函数 $f(x,y)$可由两个分量来表征:(1)入射到观察场景的光源总量和(2)场景中物体反射光的总量。相应地称为入射分量和反射分量,并分别表示为 $i(x,y)$和 $r(x,y)$。两个函数合并形成 $f(x,y)$:

$$f(x,y) = i(x,y)r(x,y) \tag{2.3.2}$$

这里

$$0 < i(x,y) < \infty \tag{2.3.3}$$

$$0 < r(x,y) < 1 \tag{2.3.4}$$

式(2.3.4)指出反射分量限制在 0(全吸收)和 1(全反射)之间。$i(x,y)$的性质取决于照射源,而 $r(x,y)$取决于成像物体的特性。应指出这种表示方式还可应用于照射光通过一个媒体形成图像的情况,如胸透 X 光片。在这种情况下,应该用透射函数代替反射函数,但其限制应该与式(2.3.4)相同,并且形成的图像函数应该如式(2.3.2)的乘积形式。

例 2.1　照射和反射的某些典型值

式(2.3.3)和式(2.3.4)给出的值是理论界限。对于可见光,下面的平均数字表明了 $i(x,y)$的典型范围。在晴朗的白天,太阳在地球表面产生的照度超过 90 000 lm/m^2。在有云的

情况下这个数值下降到 10 000 lm/m^2。在晴朗的夜晚,满月的情况下大约为 0.1 lm/m^2 的照度。类似地,下面是 $r(x,y)$的某些典型值:黑天鹅绒为 0.01,不锈钢为 0.65,白色墙为 0.80,镀银金属为 0.90,雪为 0.93。

如 2.2 节提到的,我们称单色图像在任何坐标(x_0,y_0)处的强度为图像在那一点的灰度级(l),即:

$$l = f(x_0, y_0) \tag{2.3.5}$$

很显然,从式(2.3.2)到式(2.3.4),l 位于下列范围:

$$L_{\min} \leqslant l \leqslant L_{\max} \tag{2.3.6}$$

理论上,$L_{\min}$的要求仅仅是正的,$L_{\max}$要求为有限值。实际上,$L_{\min} = i_{\min}r_{\min}$和 $L_{\max} = i_{\max}r_{\max}$。参考前面提到的平均照度和反射值范围,我们希望,$L_{\min}\approx 10$,$L_{\max}\approx 1\,000$ 作为没有附加照明的室内值的典型限制。

区间$[L_{\min}, L_{\max}]$称为灰度级。实际常常令该区间为$[0, L-1]$,这里 $l=0$ 为黑,$l=L-1$ 在灰度级中为白。所有中间值是从黑到白的各种灰色调。

2.4 图像取样和量化

从前面的讨论,我们看到有几种获取图像的方法,在所有这些方法中我们的目的是相同的,就是从感知的数据中产生数字图像。大多数传感器的输出是连续电压波形,这些波形的幅度和空间特性都与感知的物理现象有关。为了产生一幅数字图像,需要把连续的感知数据转换为数字形式。这包括两种处理:取样和量化。

2.4.1 取样和量化的基本概念

图 2.16 说明了取样和量化的基本概念。图 2.16(a)显示了一幅连续图像 $f(x,y)$,我们想把它转换为数字形式。一幅图像的 x 和 y 坐标及幅度可能都是连续的。为了把它转换为数字形式,必须在坐标和幅度上都做取样操作。数字化坐标值称为取样,数字化幅度值称为量化。

示于图 2.16(b)的一维函数是图 2.16(a)中沿线段 AB 的连续图像幅度值(灰度级)的曲线。随机变化是由图像噪声引起的。沿 AB 线段等间隔对该函数取样,如图2.16(c)所示。每个取样位置由图底部的垂直小线段给出。取样如放在函数曲线上的白色小方块所示。这一组离散位置给出了取样函数。然而,取样值在灰度级值上还是连续分布的。为了形成数字函数,灰度值也必须转变为离散值(数字化)。图 2.16(c)的右侧显示了分为 8 个离散级别的灰度级标尺,其范围从黑到白,垂直面上的小线段标记指示赋予 8 级灰度的每一级一特定值。每一取样的连续灰度值简单量化为 8 个离散灰度级中的一个。赋值是样值对小标记在垂直方向上所做的近似。由取样和量化产生的数字样值示于图 2.16(d)中。从图像的顶点开始逐行执行这一过程产生一幅二维数字图像。

按刚刚讨论的方式取样,假定有一幅在两个坐标方向和幅度上都连续的图像。实际上,取样方法由用于产生图像的传感器装置决定。如图 2.13 所示,当一幅图像由单一感知单元结合机械运动产生时,传感器的输出用前边讨论的方式量化。然而,通过选择各个机械增量的数值来完成取样,在取样处激活传感器收集数据。机械运动可以非常精确地实现,原理上,用多么

精细的程度对一幅图像取样并没有限制。然而，实际限制由在传感器上聚焦光照点的光学设备的不完善决定，光点与机械位移所能达到的精细分辨率是不一致的。

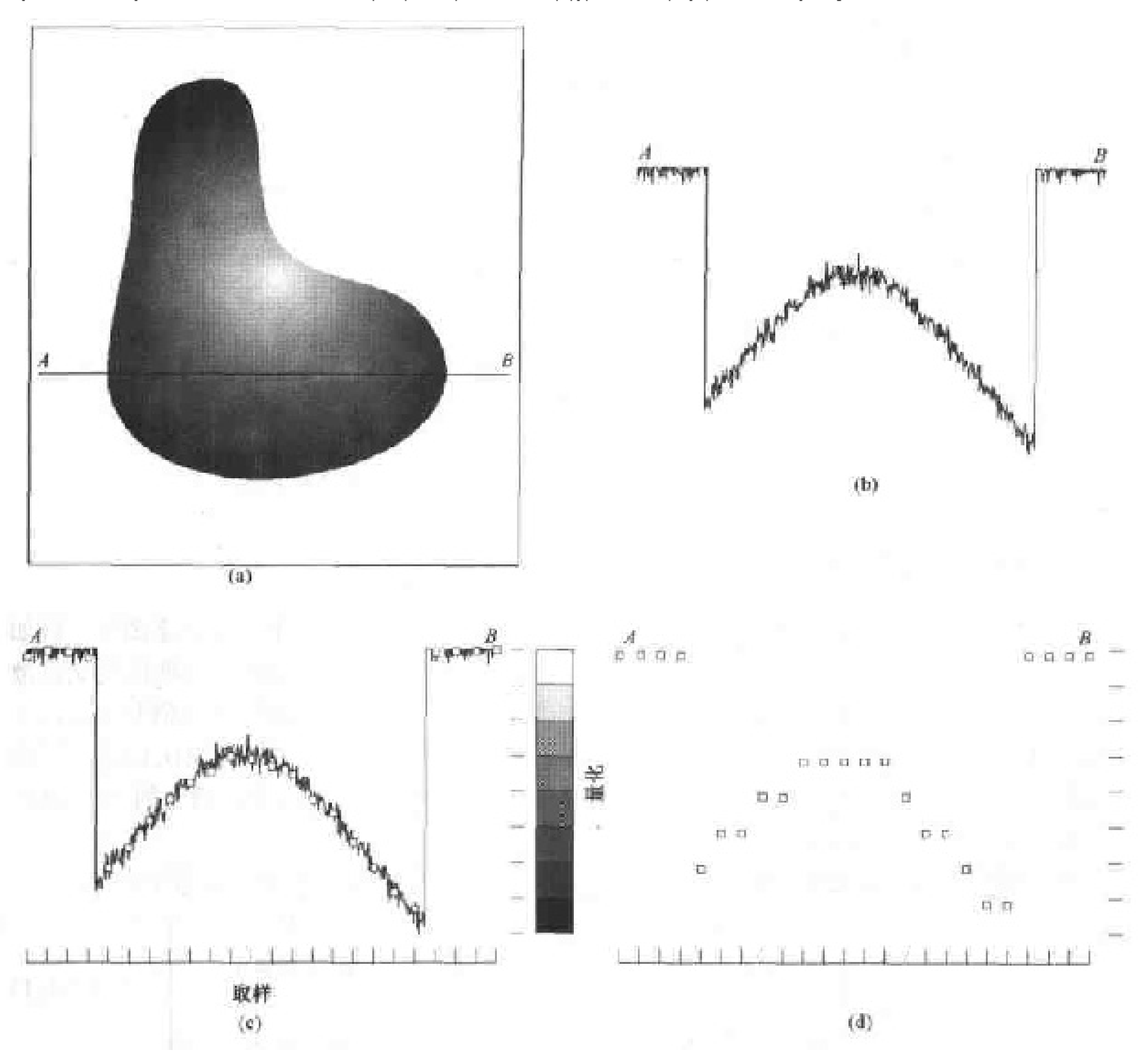


图 2.16　产生一幅数字图像。(a)连续图像，(b)在连续图像中用从 A 到 B 的扫描线来说明取样和量化的概念，(c)取样和量化，(d)数字扫描线

当用感知带获取图像时，带中传感器的数目由在一个方向上的取样限制决定。在另一个方向上的机械运动可控制得更精确一些，但是，试图在一个方向上达到某个取样密度（该密度超过了在另一个方向上由传感器数目决定的取样限制）几乎没有意义。传感器输出的量化完成了产生数字图像的过程。

当感知阵列用于图像获取时，没有运动且阵列中传感器的数量决定在两个方向上的取样限制。传感器输出的量化与前述相同，图 2.17 说明了这个概念。

图 2.17(a)显示了投影到一阵列传感器平面上的连续图像。图 2.17(b)显示了取样和量化后的图像。很明显，数字图像的质量很大程度上取决于取样和量化中所用的取样数和灰度级。然而，正如 2.4.3 节显示的那样，在选择这些参数时，图像内容是一个重要考虑因素。

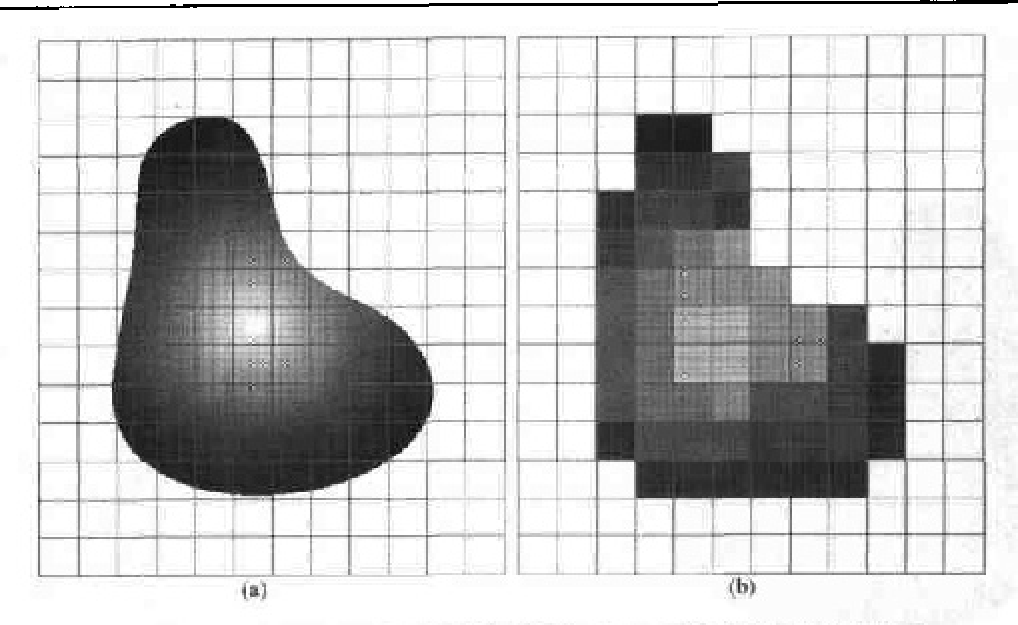


图 2.17 (a)连续图像投影到传感器阵列,(b)图像取样和量化的结果

2.4.2 数字图像表示

取样和量化的结果是一个实际矩阵。在本书中将用两种主要方法来表示数字图像。假如一幅图像 $f(x,y)$被取样,则产生的数字图像有 M 行和 N 列。现在,坐标(x,y)的值变成离散量。为表达清楚和方便起见,对这些离散坐标应该用整数。这样,原点的坐标值是$(x,y)=(0,0)$。沿图像第一行的下一个坐标值用$(x,y)=(0,1)$来表示。记住,用符号$(0,1)$表示沿第一行的第二个取样是很重要的,但它并不意味着是图像取样的物理坐标实际值。图 2.18 显示了贯穿本书所用的坐标约定。

前一节介绍的表示法允许我们以下面的紧凑矩阵形式写出完整的 $M\times N$ 数字图像:

$$f(x,y)=\begin{bmatrix} f(0,0) & f(0,1) & \cdots & f(0,N-1) \\ f(1,0) & f(1,1) & \cdots & f(1,N-1) \\ \vdots & & & \\ f(M-1,0) & f(M-1,1) & \cdots & f(M-1,N-1) \end{bmatrix} \tag{2.4.1}$$

这个表达式的右侧定义了一幅数字图像。矩阵中的每个元素称为图像单元、图像元素或像素,图像和像素这两个术语将在我们讨论的其余部分用以表示数字图像和它的元素。

在某些讨论中,可以用传统矩阵表示法来表示数字图像和像素:

$$\mathbf{A}=\begin{bmatrix} a_{0,0} & a_{0,1} & \cdots & a_{0,N-1} \\ a_{1,0} & a_{1,1} & \cdots & a_{1,N-1} \\ \vdots & \vdots & & \vdots \\ a_{M-1,0} & a_{M-1,1} & \cdots & a_{M-1,N-1} \end{bmatrix} \tag{2.4.2}$$

显然,$a_{ij}=f(x=i,y=j)=f(i,j)$,因此,式(2.4.1)和式(2.4.2)是恒等矩阵。

以更正规的数学术语描述取样和量化有时可能有用。令 Z 和 R 分别表示整数集和实数集。取样过程可看做是把 xy 平面分为网格,每一网格中心的坐标是笛卡儿积 Z^2 的一对元素,Z^2 是所有有序元素对(z_i,z_j)的集合,z_i 和 z_j 是 Z 中的整数。因此,如果(x,y)是 Z^2 中的

整数，并且 f 是把灰度级值(即实数集 R 的一个实数)赋予每个特定坐标对 (x, y) 的函数，则 $f(x, y)$ 就表示一幅数字图像。很显然，这个函数赋值过程就是前边描述的量化过程。如果灰度也是整数(在本章的后续章节通常是这种情况)，则 Z 代替 R，数字图像变成一个二维函数且它的坐标和幅值都是整数。

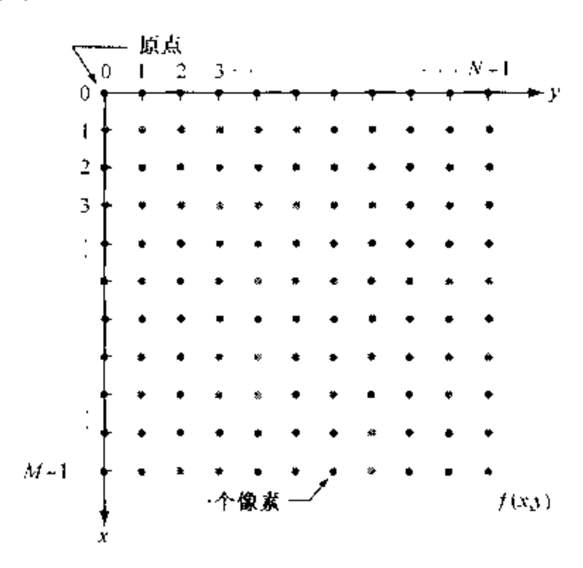


图 2.18　本书表示数字图像所用的坐标约定

数字化过程对于 M、N 值和每个像素允许的离散灰度级数 L 需要一个判定。对 M 和 N 除了必须取正整数外没有其他要求。然而，出于处理、存储和取样硬件的考虑，灰度级典型的取值是 2 的整数次幂：

$$L = 2^k \tag{2.4.3}$$

这里，假设离散灰度级是等间隔的并且是区间 $[0, L-1]$ 内的整数。有时灰度级取值范围称为图像的动态范围。把占有灰度级全部有效段的图像叫做高动态范围图像。当相当可观数目的像素呈现这样的特性时，图像就有较高对比度。相反，低动态范围的图像看上去似乎是冲淡了的灰暗格调。这一点在 3.3 节将更加详细地讨论。

数字 b 是存储数字图像需要的比特数，有：

$$b = M \times N \times k \tag{2.4.4}$$

当 $M = N$ 时，上式变为：

$$b = N^2 k \tag{2.4.5}$$

表 2.1 显示了，对于 N 和 k 的不同取值，需要用来存储方形图像的比特数。括号中显示的是相对每一 k 值的灰度级数。

当一幅图像有 2^k 灰度级时，实际上通常称该图像是 k 比特图像。例如，一幅图像有 256 个可能的灰度级，称其为 8 比特图像。注意，大小为 1 024 × 1 024 的 8 比特图像的存储是必要的，并且再高就没有意义了。

表 2.1 N 和 K 变化时存储的比特数

N/k	1($L=2$)	2($L=4$)	3($L=8$)	4($L=16$)	5($L=32$)	6($L=64$)	7($L=128$)	8($L=256$)
32	1 024	2 048	3 072	4 096	5 120	6 144	7 168	8 192
64	4 096	8 192	12 288	16 384	20 480	24 576	28 672	32 768
128	16 384	32 768	49 152	65 536	81 920	98 304	114 688	131 072
256	65 536	131 072	196 608	262 144	372 680	393 216	458 752	524 288
512	262 144	524 288	786 432	1 048 576	1 310 720	1 572 864	1 835 008	2 097 152
1 024	1 048 576	2 097 152	3 145 728	4 194 304	5 242 880	6 291 456	7 340 032	8 388 608
2 048	4 194 304	8 388 608	12 582 912	16 777 216	20 971 520	25 165 824	29 369 128	33 554 432
4 096	16 777 216	33 554 432	50 331 648	67 108 864	83 886 080	100 663 296	117 440 512	134 217 728
8 192	67 108 864	134 217 728	201 326 592	268 435 456	335 544 320	402 653 184	469 762 048	536 870 912

2.4.3 空间和灰度级分辨率

取样值是决定一幅图像空间分辨率的主要参数。基本上,空间分辨率是图像中可辨别的最小细节。假定我们画一幅宽度为 W 的垂直线的图,在线间还有宽度为 W 的线。线对是由一条线与它紧邻的线组成的。这样,线对的宽度为 $2W$,并且每单位距离有 $1/2W$ 对线。广泛使用的分辨率的意义是在每单位距离可分辨的最小线对数目,例如,每毫米 100 线对。

类似地,灰度级分辨率是指在灰度级别中可分辨的最小变化。但是,正像 2.1.3 节提到的那样,在灰度级中,测量可分辨的变化是一个高度主观的过程。这里考虑了用以产生数字图像取样数目的判断方法,但是,对于灰度级数,这种方法却不可行。正如前面章节提到的,由于硬件方面的考虑,灰度级数通常是 2 的整数次幂。大多数情况该值取 8 比特,在某些特殊的灰度增强的应用场合可能用 16 比特。有时,我们寻求以 10 或 12 比特精度数字化一幅图像的系统,但这些系统都是特例而不是常规系统。

当没有必要对涉及像素的物理分辨率进行实际度量和在原始场景中分析细节等级时,通常就把大小为 $M\times N$,灰度为 L 级的数字图像称为空间分辨率为 $M\times N$ 像素、灰度级分辨率为 L 级的数字图像。为清楚起见,在后续的讨论中我们将经常使用这一术语,仅在需要时参照实际分辨率细节。

例 2.2 在数字图像中各种取样数目的典型效果

图 2.19 显示了一幅 1 024×1 024 像素的图像,其灰度用 8 比特表示。示于图 2.19 的其他图像是对 1 024×1 024 图像抽样的结果。抽样是从原图像中删去一样的行和列数完成的。例如,512×512 图像是从 1 024×1 024 图像中每隔一行或一列删去一行或一列得到的。256×256 图像是从 512×512 图像中每隔一行或一列删去一行或一列产生的,等等。其灰度级保持在 256。

这些图像显示了不同取样密度间的大小比例,但是,它们大小的差别使得很难看出减少取样数目带来的影响。比较这一效果的最简单办法是,通过复制行或列使抽样后的图像仍复原到 1 024×1 024 大小。其结果示于图 2.20(b)到(f)中。图 2.20(a)是示于图 2.19 的 1 024×1 024,256 级灰度的图像,把它复制到这里以便于比较。

用图 2.20(b)中的 512×512 图像与图 2.20(a)比较,并注意,事实上不可能说出这两幅图像的区别。细节损失程度太细微以至于不可能看出这些图像在这一尺度上印刷出来后的区别。下一个,图 2.20(c)中的 256×256 图像,在花瓣和黑色背景间的边缘上显示了细微的棋盘格

图案,在整个图像上稍微明显的颗粒也开始出现。这些效果在图 2.20(d)128×128 的图像中很明显,并且在图 2.20(e)和(f)的 64×64 和 32×32 的图像中显得更加明显了。

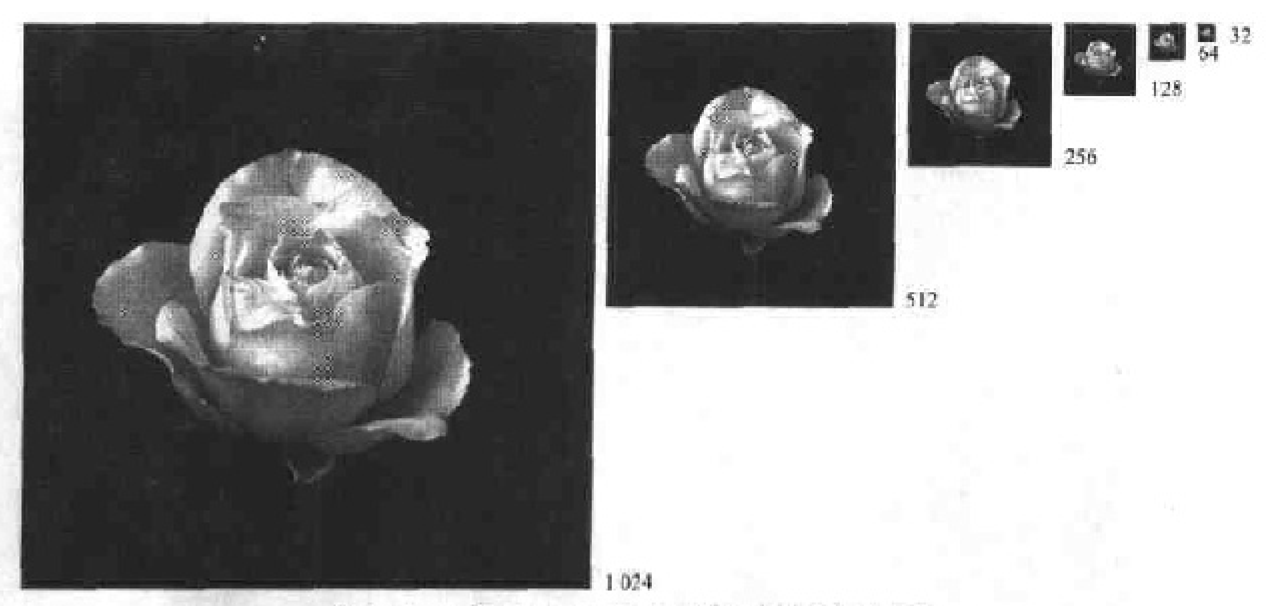


图 2.19　一幅 1 024×1 024,8 比特的图像逐次减少取样至32×32像素,可允许的灰度级保持为256

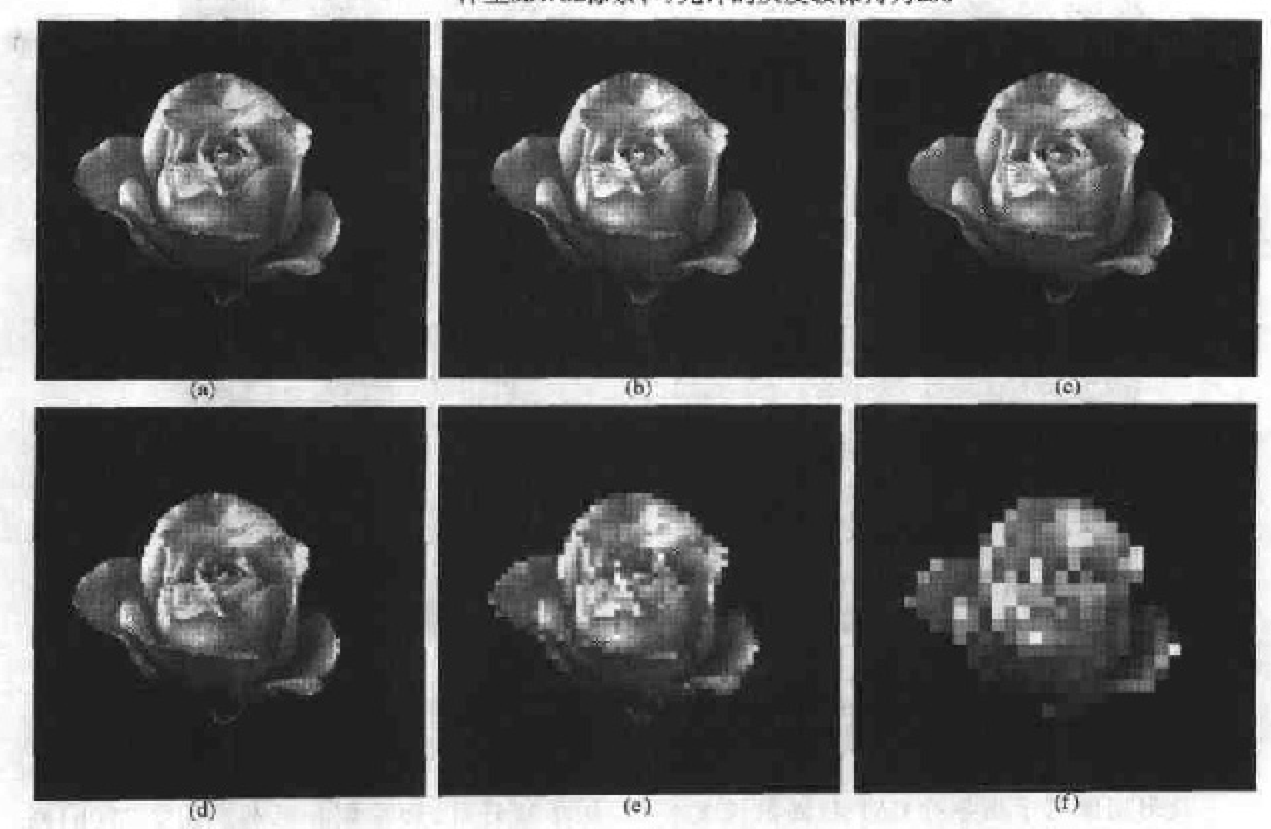


图 2.20　(a)1 024×1 024,8 比特图像,(b)用复制行或列的方法把 512×512 图像复原为 1 024×1 024 像素的图像,(c)到(f)256×256,128×128,64×64 和 32×32 图像复原到 1 024×1 024 像素的图像

例 2.3 在数字图像中灰度级值变化的典型效果

在这个例子中,我们保持取样数恒定而以 2 的整数次幂方式把灰度级从 256 减少到 2。图 2.21(a)是一幅大小为 452×374 的 CAT 投影图像,它以 $k=8$(256 灰度级)显示。像这样的图像是由将 X 射线源固定在一个位置得到的,这样产生了一幅在任何指定方向上的二维图像。

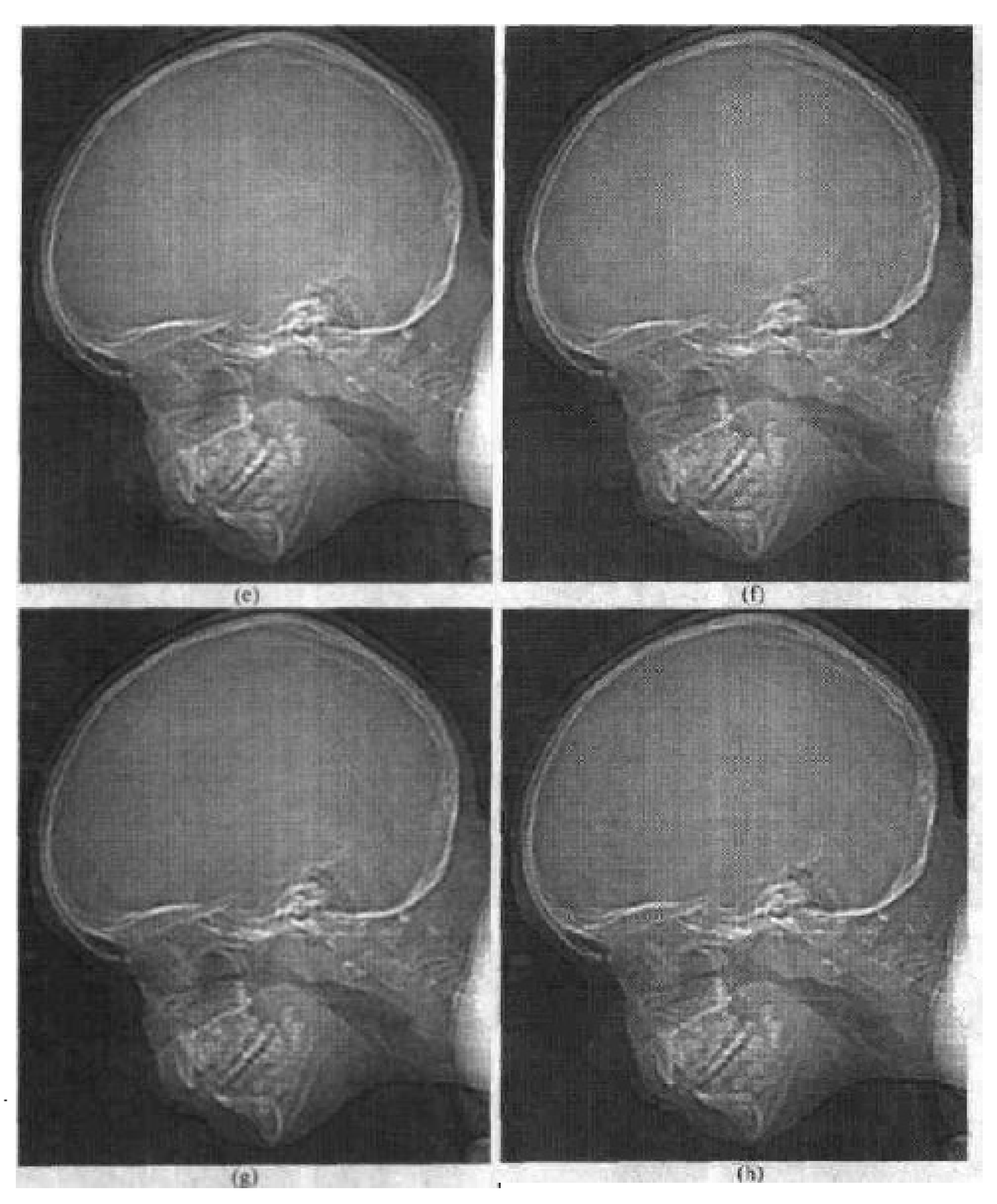


图 2.21 (a)453×374,256 灰度级图像;(b)~(d)保持空间分辨率为常数,以灰度级为128,64和32显示的图像

投影图像用于指导对 CAT 扫描器设置参数,包括倾斜角、切片数和范围。图 2.21(b)到(h)是在保持空间分辨率 452×374 像素不变化的情况下,比特数从 $k=7$ 减到 $k=1$ 得到的图像。256,128 和 64 灰度级图像的全部实际效果都是视觉上可分辨的。然而,示于图 2.21(d)的 32 级灰度的图像在平滑灰度区域内(特别在头盖骨处)有一组几乎看不见的非常细小的山脊状结构。这种效果是由在数字图像平滑区灰度级数量不足引起的,它被

称为伪轮廓,之所以这样叫是因为山脊像地图中的地形等值线。伪轮廓通常在均匀的平面上以 16 级或更少灰度显示的图像中十分明显,如图 2.21(e)到(h)所示的图像。

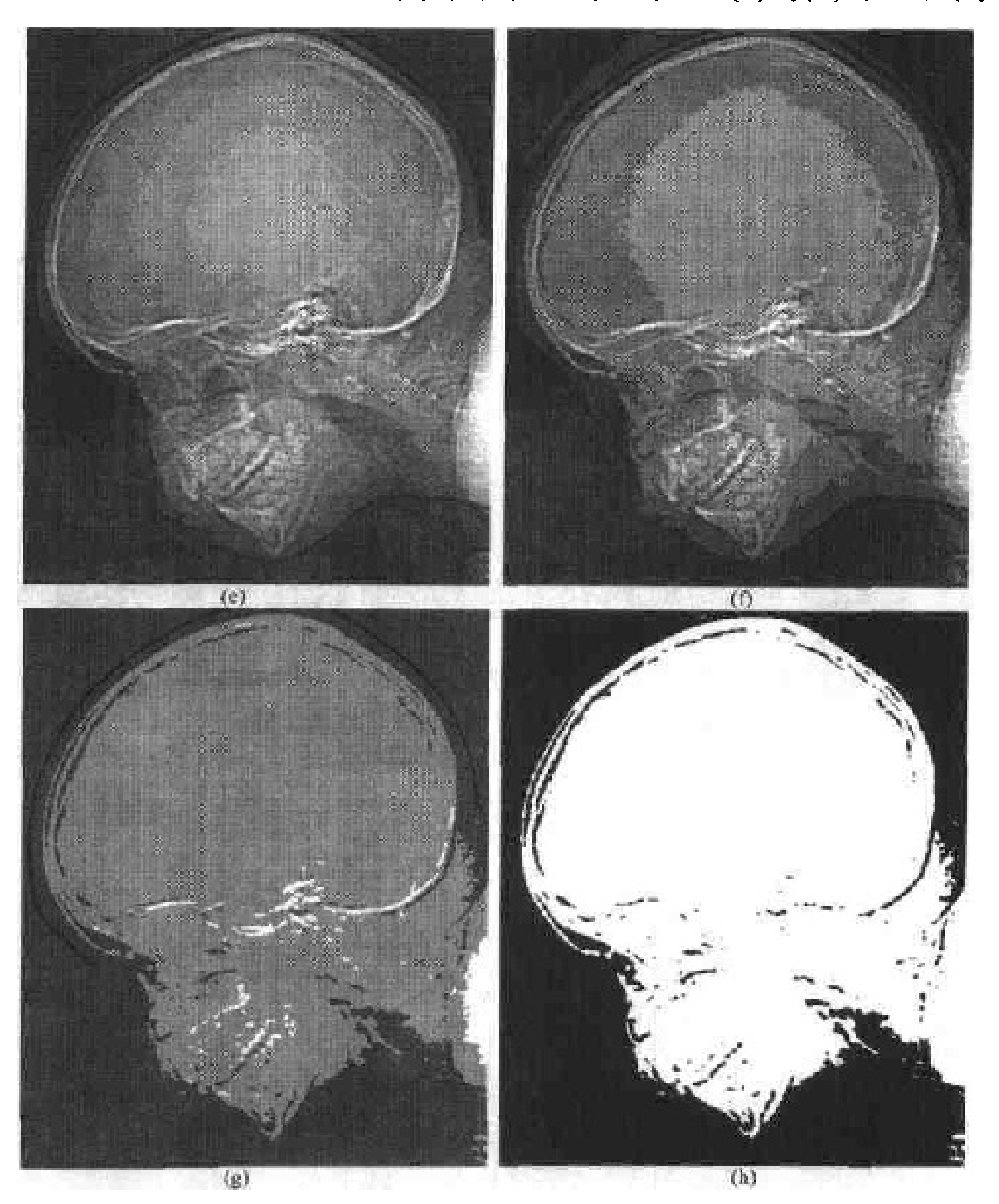


图 2.21(续)　(e)~(h)以 16,8,4 和 2 灰度级显示的图像(原图像由 Vanderbilt 大学法学中心放射学和放射科学系的David R.Pickens博士提供)

作为一个大概的缩略图规则(并且为方便起见),假设数值为 2 的整数幂、256×256 像素大小、64 级灰度的图像一般是可以接受的,该图像是没有明显可见的取样棋盘格和伪轮廓的最小图像。

例 2.2 和例 2.3 的结果表明了分别改变 N 和 k 在图像质量方面产生的结果。然而,这些结果也仅仅部分地回答了改变 N 和 k 如何影响图像的问题,因为远没有考虑这两个参数间可能存在的关系。Huang[1965]早期的研究试图通过实验,用数量来表示同时改变 N 和 k 对图像质量所产生的影响。实验由一组主观测试组成,采用类似图 2.22 所示的图像。妇女脸庞图像

含有较少细节,摄像师图像包含中等程度的细节,人群图像相比之下包含大量的细节。

这一组图像由改变 N 和 k 产生,然后要求观察者将图像按主观判定质量排序。以在 Nk 平面上的等偏爱曲线形式总结这一结果(图 2.23 显示了示于图 2.22 图像的平均等偏爱曲线表示),Nk 平面中每一点表示一幅图像,该图像的 N 和 k 值等于该点的坐标。位于等偏爱曲线上的点对应图像的主观判定质量。在实验中发现等偏爱曲线向右上方移动,但是这三类图像的等偏爱曲线形状与示于图 2.23 的那些曲线类似。这是所希望的,因为在曲线中,简单地向上、向右移动意味着 N 和 k 的值变大,图像质量好。

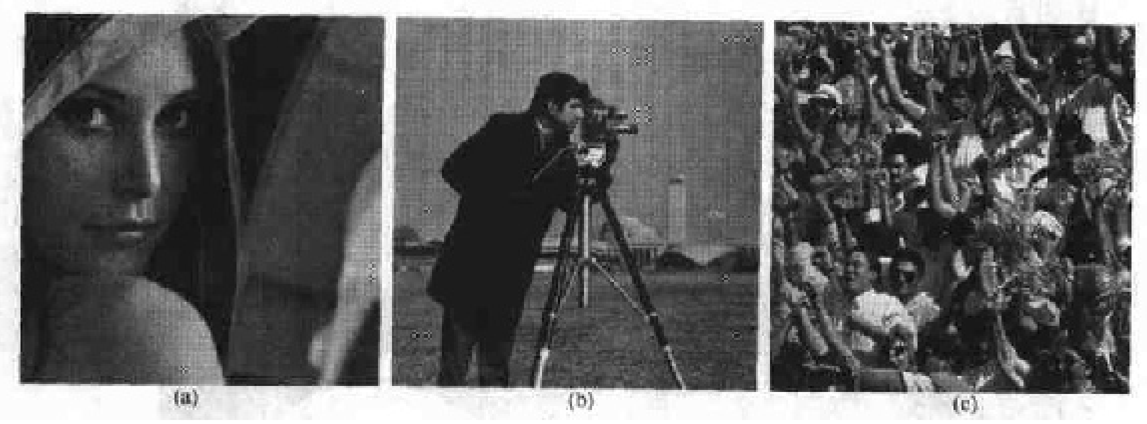


图 2.22 (a)具有少量细节的图像,(b)具有中等细节的图像,(c)具有大量细节的图像[图(b)由麻省理工学院提供]

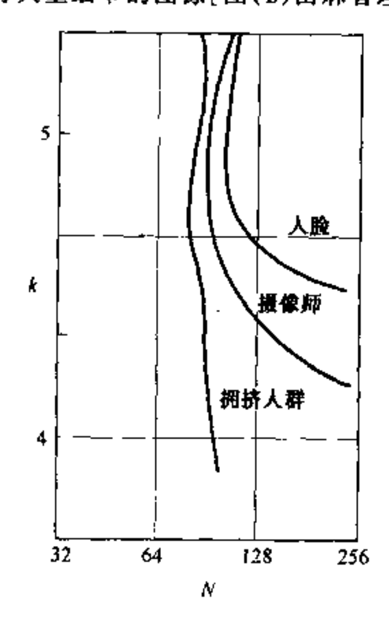


图 2.23 图 2.22 中三类图像的典型等偏爱曲线

当前讨论的关键点是当图像细节增加时等偏爱曲线变得更垂直。这个结果表明对于有大量细节的图像可能只需要少数的灰度级。例如,图 2.23 中与人群图像相对应的等偏爱曲线近于垂直。这里指出,N 值固定时,这类图像的感觉质量与所用的灰度级数(对于图 2.23 所示的灰度级范围)是近似独立的。更有意思的是,我们注意到,另两类图像的感觉质量在空间分辨率增加的情况下,于某一段间隔内保持相同,但是,灰度级数目实际上在减少。对于这一结果

最可能的原因是，k 的减少导致图像对比度的明显增加，产生图像质量改进后人通常感受到的视觉效果。

2.4.4　混淆的水纹图样

如第 4 章讨论的那样，曲线之下的区域是有限的那些函数可以用频率可变的正弦或余弦函数来表示。高频的正弦/余弦分量决定函数的最高频率含量。假定，这个最高频率是受限的，并且该函数时间上是无限的(这些函数称为频带受限函数)。香农取样定理[Bracewell (1995)]告诉我们，如果一个函数以等于或大于其最高频率的 2 倍取样，那么就能从取样完全复原原函数。如果函数被亚取样，则混淆的现象将恶化取样的图像。混淆以附加频率分量的形式被引入取样后的函数。这些频率称为混淆频率。注意，图像的取样率是单位距离的取样数目(在两个空间方向上)。

如证明的那样，除了在后面章节讨论的特殊情况外，实际上满足取样定理是不可能的。我们只能用时间有限的取样数据去工作。可以仿照用一个选通函数去乘无限函数的方法把时间无限的函数转换为时间有限的函数，选通函数在某一时间间隔取值为 1，其他时为 0。遗憾的是，选通函数本身也有无限扩展的频率分量。这样，在对频带受限函数限制时间时导致限制频带失效，这将使其违背取样定理的关键条件。在图像中，消除混淆效应的主要方法是在取样前通过模糊图像(在第 4 章将详细讨论模糊问题)降低引起混淆的高频分量。然而，在取样图像中混淆总是存在的。混淆频率的影响在正确条件下可看成是以下边讨论的叫做"水纹图样"的形式。

有一种重要的特殊情况，在这种情况下，在有限的时间间隔内不违反取样定理的条件下，对时间无限的函数取样。当一个函数是周期函数时，可以按等于或超过其最高频率 2 倍的频率对它取样，在函数周期整数倍的准确位置捕获样点时，从样点复原该函数是可能的。这个特殊情况使我们可以生动地说明水纹效应。图 2.24 显示了两个周期相同的等间隔垂直条纹的图样，以相反的方向旋转，然后互相重叠，使两幅图像相乘。在图 2.24 中可以看到周期性断裂导致的水纹图样，作为在垂直方向走的二维正弦(混淆的)波形(看上去像波纹状的白铁皮屋顶)。由周期性墨点组成的印刷品数字化(例如扫描)图像时，也会出现类似的图样。

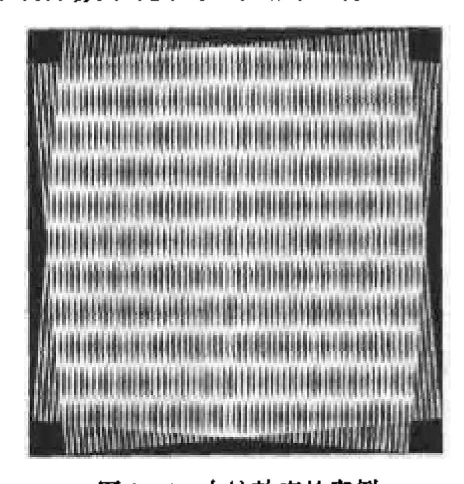


图 2.24　水纹效应的实例

2.4.5 放大和收缩数字图像

用如何放大和收缩数字图像的简单讨论来结束取样和量化的论述。这个题目关系到取样和量化,因为放大可看做过采样,收缩可看做欠采样。这两种操作与取样和量化一幅原始连续图像之间的关键区别是放大和收缩适用于数字图像。

放大要求执行两步操作:创立新的像素位置和对这些新位置赋灰度值。让我们用一个简单的例子开始。假定,有一幅大小为 500×500 像素的图像,想把它扩大为 1.5 倍,即 750×750 像素。概念上看,一种最容易的形象化放大方法是在原始图像上放一个虚构的 750×750 的栅格。很显然,栅格的间隔应小于 1 个像素,因为,我们在一个较小的图像上去拟合它。为了对覆盖层上的任何点进行灰度赋值,我们在原图像上寻找最靠近的像素并把它的灰度值赋给栅格上的新像素。当对覆盖栅格的全部点都赋完值后,简单地把它扩展到原来指定的大小,以得到放大的图像。这种灰度赋值法叫做最近邻域内插(像素相邻在下节讨论)。

产生图 2.20(b)到(f)图像所用的像素内插法是最近邻域内插的一个特殊情况。当想要以整数倍增加图像的大小时,可采用像素复制的方法。例如,把图像增大一倍,可以复制每一列。这就使图像在水平方向增加一倍。然后复制已增大了的图像的每一行以使图像在垂直方向上增大一倍。使用相同的方法可以任何整数倍增大图像(3 倍、4 倍等)。复制仅仅是为了达到希望的大小以要求的倍数去操作。每个像素的灰度赋值过程由这样的事实决定,即新位置的值准确复制旧位置的值。虽然最近邻域插值较快,但它有不希望的特性,即它会产生棋盘格效应,在高放大倍数时这一效应特别明显。图 2.20(e)和(f)是这一问题的典型例子。一种稍有改进的灰度赋值方法是采用 4 个最邻近点的双线性内插。令(x',y')代表放大图像中一点的坐标(把它想像为前边讨论的栅格上的点),并令 $v(x',y')$代表它被赋予的灰度值。对双线性内插来说,所赋灰度值由下式给出:

$$v(x',y') = ax' + by' + cx'y' + d \tag{2.4.6}$$

这里,4 个系数由点(x',y')的 4 个最近邻点写出的 4 个未知方程决定。

图像收缩可以与刚刚讨论过的放大相类似的方法操作。像素复制的等效过程是行-列删除。例如,把图像缩小,可以每隔一行(或一列)删除一行(或一列)。可以用放大栅格模拟用非整数参数收缩,设想用扩大的栅格在原图像上拟合,做灰度级最近邻域或双线性内插,然后把栅格收缩到原来规定的大小。为了减小可能的混淆效应,可以在收缩之前稍微模糊一下图像。数字图像的模糊在第 3 章和第 4 章中讨论。

还可以使用更多邻点的内插。使用更多邻点意味着在更复杂表面对点赋灰度值,该方法通常会产生平滑一些的结果。对于三维图形[Watt(1993)]和医学图像处理[Lehmann 等(1999)],这是一个特别重要的产生图像的方法。但是,对于通常目的的数字图像放大和收缩,减少一些额外的计算负担是合理的,因此双线性内插通常是首选方法。

例 2.4 用双线性内插放大图像

图 2.25 的上面一行显示了图 2.20(d)到(f),这些图像是用最近邻域内插法,从 128×128,64×64 和 32×32 放大到 1 024×1 024 像素的。用双线性内插方法的结果示于图 2.25 的第二行。总的看来改进是很明显的,特别对于 128×128 和 64×64 的情况。32×32 到1 024×1 024图像的放大结果是模糊的,但是请记住,该图是用 32 这一参数放大的。尽管如此,示于图 2.25(f)的双线性内插结果还是可接受的原图像形状的再现,图 2.25(c)所示的是较失败的一幅。

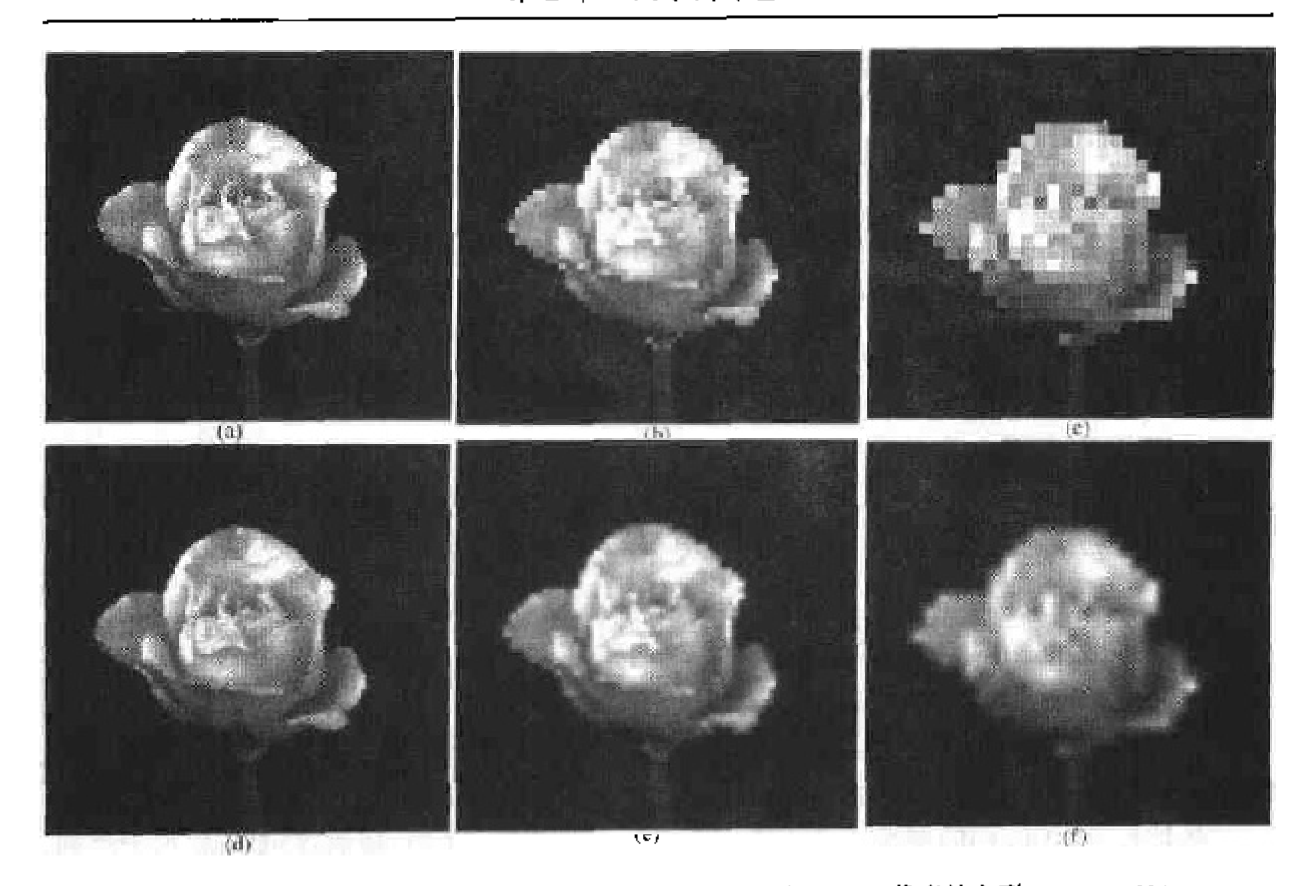


图 2.25　上一行：用最近邻域内插法从 128 × 128，64 × 64 和 32 × 32 像素放大到 1 024 × 1 024 像素的图像。下一行：按相同的参数顺序，用双线性内插法得到的放大图像

2.5　像素间的一些基本关系

在这一节，考虑在数字图像中像素间的一些重要关系。如前边提到的那样，一幅图像用 $f(x,y)$ 表示。当我们指特殊像素时用小写字母（如 p 和 q）表示。

2.5.1　相邻像素

位于坐标 (x,y) 的一个像素 p 有 4 个水平和垂直的相邻像素，其坐标由下式给出：

$$(x+1,y),(x-1,y),(x,y+1),(x,y-1)$$

这个像素集称为 p 的 4 邻域，用 $N_4(p)$ 表示。每个像素距 (x,y) 一个单位距离，如果 (x,y) 位于图像的边界，则 p 的某一邻像素位于数字图像的外部。

p 的 4 个对角邻像素有如下坐标：

$$(x+1,y+1),(x+1,y-1),(x-1,y+1),(x-1,y-1)$$

并用 $N_D(p)$ 表示。与 4 个邻域点一起把这些点叫做 p 的 8 邻域，用 $N_8(p)$ 表示。与前边一样，如果 (x,y) 位于图像的边界，则 $N_D(p)$ 和 $N_8(p)$ 中的某些点落入图像的外边。

2.5.2　邻接性、连通性、区域和边界

像素间的连通性是一个基本概念，它简化了许多数字图像概念的定义，如区域和边界。为

了确定两个像素是否连通,必须确定它们是否相邻及它们的灰度值是否满足特定的相似性准则(或者说,它们的灰度值是否相等)。例如,在具有 0,1 值的二值图像中,两个像素可能是 4 邻接的,但是,仅仅当它们仅仅具有同一灰度值时才能说是连通的。

令 V 是用于定义邻接性的灰度值集合。在二值图像中,如果把具有 1 值的像素归入邻接的,则 $V=\{1\}$。在灰度图像中,概念是一样的,但是集合 V 一般包含更多的元素。例如,对于具有可能的灰度值且在 0 到 255 范围内的像素邻接性,集合 V 可能是这 256 个值的任何一个子集。考虑三种类型的邻接性:

(a)**4 邻接**:如果 q 在 $N_4(p)$ 集中,具有 V 中数值的两个像素 p 和 q 是 4 邻接的。

(b)**8 邻接**:如果 q 在 $N_8(p)$ 集中,则具有 V 中数值的两个像素 p 和 q 是 8 邻接的。

(c)**m 邻接(混合邻接)**:如果(i) q 在 $N_4(p)$ 中,或者(ii) q 在 $N_D(p)$ 中且集合 $N_4(p)\cap N_4(q)$ 没有 V 值的像素,则具有 V 值的像素 p 和 q 是 m 邻接的。

混合邻接是 8 邻接的改进。混合邻接的引入是为了消除采用 8 邻接常常发生的二义性。例如,考虑图 2.26(a)对于 $V=\{1\}$所示的像素安排。位于图 2.26(b)上部的三个像素显示了多重(二义性)8 邻接,如虚线指出的那样。这种二义性可以通过 m 邻接消除,如图 2.26(c)所示。如果 S_1 中的某些像素与 S_2 中的某些像素邻接,则两个图像子集 S_1 和 S_2 是相邻接的。这里和下面的定义中邻接意味着 4、8 或者 m 邻接。

从具有坐标(x,y)的像素 p 到具有坐标(s,t)的像素 q 的通路(或曲线)是特定像素序列,其坐标为:

$$(x_0,y_0),(x_1,y_1),\cdots,(x_n,y_n)$$

这里$(x_0,y_0)=(x,y)$,$(x_n,y_n)=(s,t)$,并且像素(x_i,y_i)和(x_{i-1},y_{i-1})(对于 $1\leqslant i\leqslant n$)是邻接的。在这种情况下,n 是通路的长度。如果$(x_0,y_0)=(x_n,y_n)$,则通路是闭合通路。可以依据特定的邻接类型定义 4,8 或 m 邻接。例如,图 2.26(b)所示的,东北角和东南角点之间的通路是 8 通路,图 2.26(c)中的通路是 m 通路。注意在 m 通路中不存在二义性。

令 S 代表一幅图像中像素的子集。如果在 S 中全部像素之间存在一个通路,则可以说两个像素 p 和 q 在 S 中是连通的。对于 S 中的任何像素 p,S 中连通到该像素的像素集叫做 S 的连通分量。如果 S 仅有一个连通分量,则集合 S 叫做连通集。

令 R 是图像中的像素子集。如果 R 是连通集,则称 R 为一个区域。一个区域 R 的边界(也称为边缘或轮廓)是区域中像素的集合,该区域有一个或多个不在 R 中的邻点。如果 R 是整幅图像(我们设这幅图像是像素的方形集合),则边界由图像第一行、第一列和最后一行一列定义。这个附加定义是需要的,因为图像除了边缘没有邻点。正常情况下,当我们提到一个区域时,指的是一幅图像的子集,并且区域边界中的任何像素(与图像边缘吻合)都作为区域边界部分全部包含于其中。

边缘的概念在涉及区域和边界的讨论中常常遇到。然而,这些概念中有一个关键区别。一个有限区域的边界形成一条闭合通路(习题 2.14),并且是"整体"概念。正如第 10 章详细讨论的,边缘是由具有某些导数值(超过预先设定的阈值)的像素形成。这样,边缘的概念是基于在不连续点进行灰度级测量的局部概念。把边缘点连接成边缘线段是可能的,并且有时以与边界对应的方法连接线段,但并不总是这样。边缘和边界吻合的一个例外就是二值图像的情况。根据连通类型和所用的边缘算子(在第 10 章中讨论),从二值区域提取边缘与提取区域边界是一样

的,这很直观。在概念上,把边缘考虑为强度不连续的点和封闭通路的边界是有帮助的。

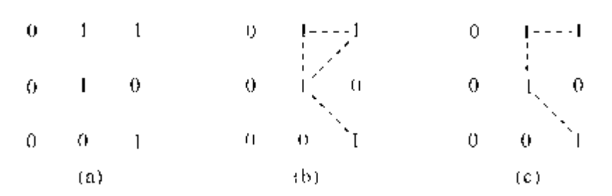


图 2.26　(a)像素安排,(b)中心像素的 8 邻接像素(虚线所示),(c)m 邻接

2.5.3 距离度量

对于像素 p,q 和 z,其坐标分别为 (x,y),(s,t) 和 (v,w),如果:

(a) $D(p,q)\geqslant 0$ $[D(p,q)=0$,当且仅当 $p=q]$

(b) $D(p,q)=D(q,p)$

(c) $D(p,z)\leqslant D(p,q)+D(q,z)$

则 D 是距离函数或度量。

p 和 q 间的欧氏距离定义于下:

$$D_e(p,q)=[(x-s)^2+(y-t)^2]^{\frac{1}{2}} \tag{2.5.1}$$

对于距离度量,距点 (x,y) 的距离小于或等于某一值 r 的像素是,中心在 (x,y) 半径为 r 的圆平面。

p 和 q 间的距离 D_4(也叫城市街区距离)如下式定义:

$$D_4(p,q)=|x-s|+|y-t| \tag{2.5.2}$$

在这种情况下,距 (x,y) 的 D_4 距离小于或等于某一值 r 的像素形成一个中心在 (x,y) 的菱形。例如,距 (x,y) 的 D_4 距离小于或等于 2 的像素形成固定距离的下列轮廓:

```
    2
  2 1 2
2 1 0 1 2
  2 1 2
    2
```

具有 $D_4=1$ 的像素是 (x,y) 的 4 邻域。

p 和 q 间的 D_8 距离(也叫棋盘距离)定义为下式:

$$D_8(p,q)=\max(|x-s|,|y-t|) \tag{2.5.3}$$

在这种情况下,距 (x,y) 的 D_8 距离小于或等于某一值 r 的像素形成中心在 (x,y) 的方形。例如,距点 (x,y)(中心点)的 D_8 距离小于或等于 2 的像素形成下列固定距离的轮廓:

```
2 2 2 2 2
2 1 1 1 2
2 1 0 1 2
2 1 1 1 2
2 2 2 2 2
```

具有 $D_8=1$ 的像素是关于(x,y)的8邻域。

注意,p 和 q 之间的 D_4 和 D_8 距离与任何通路无关,通路可能存在于各点之间,因为这些距离仅与点的坐标有关。然而,如果选择考虑 m 邻接,则两点间的 D_m 距离用点间最短的通路定义。在这种情况下,两像素间的距离将依赖于沿通路的像素值以及它们的邻点值。例如,考虑下列安排的像素并假设 p,p_2 和 p_4 的值为1,p_1 和 p_3 的值为0或1:

$$\begin{matrix} & p_3 & p_4 \\ p_1 & p_2 & \\ p & & \end{matrix}$$

假设考虑值为1的像素邻接(即:$V=\{1\}$)。如果 p_1 和 p_3 是0,p 和 p_4 间最短 m 通路的长度(D_m 距离)是2。如果 p_1 是1,则 p_2 和 p 将不再是 m 邻接(见 m 邻接的定义),并且 m 通路的长度变为3(通路通过点 $p\,p_1\,p_2\,p_4$)。类似地,如果 p_3 是1(并且 p_1 为0),则最短的通路距离也是3。最后,p_1 和 p_3 都为1,则 p 和 p_4 间的最短 m 通路长度为4,在这种情况,通路通过点 $p\,p_1\,p_2\,p_3\,p_4$。

2.5.4 基于像素的图像操作

在后续各章,大量的文献是关于像素间运算的,例如,用一幅图像去除另一幅图像。在式(2.4.2)中,图像以矩阵的形式表示,而矩阵除法无定义。因此当我们提到类似用一幅图像除另一幅图像的运算时,意思是在两幅图像相应的像素间执行除法运算。例如,如果 f 和 g 是两幅图像,用 g 除 f 形成的图像的第一个像素值是用 g 中的第一个像素去除 f 中的第一个像素的结果,当然,是假设在 g 中没有一个像素值为0。其他的算术和逻辑操作也类似地定义为图像中相对应像素间的操作。

2.6 线性和非线性操作

令 H 是一种算子,其输入和输出都是图像。如果对于任何两幅图像 f 和 g 及任何两个标量 a 和 b 有如下关系,则称 H 为线性算子:

$$H(af+bg)=aH(f)+bH(g) \tag{2.6.1}$$

换句话说,对两幅图像(用所示的常数去乘)的和应用线性算子等同于分别对图像应用该算子,并各自与适当的常数相乘,然后将结果相加。例如,对 K 幅图像求和的算子是一个线性算子。计算两幅图像差分绝对值的算子就不是线性算子。不能通过式(2.6.1)检验的算子就定义为非线性算子。

线性算子在图像处理中特别重要。因为它们是充分了解理论和实践结果的主要基础。虽然非线性算子也会提供较好的性能,但是它们不是总可以预测的,大部分不能在理论上很好地理解。

小结

本章的材料对后续的讨论提供了基本背景信息。对人类视觉系统的论述虽然简单,但提供了人眼感知图像信息能力的一个基本概念。光和电磁波的讨论是理解本书所用的许多图像

来源的基础。类似地,在 2.3.4 节介绍的图像模型作为在第 4 章中称为"固态滤波"图像增强技术的基础,并且在第 10 章中用来解释图像直方图形状的影响。

2.4 节介绍的取样概念是实践中所遇到的许多现象的基础。这些概念可以进一步扩展为掌握和了解频率内容的基础，频率域更详细的讨论在第 4 章中给出。取样混淆效果的概念在图像获取中也是很重要的。

2.5 节介绍的概念是基于像素邻域的处理技术的基本组成部分。如下一章和第 5 章所示,邻域处理方法是许多增强和复原过程的核心。在应用时,邻域处理由于处理速度和软硬件实现简单,很适合图像处理的商业应用。最后,线性算子的概念和与其相关的理论和计算功能将在后续三章中广泛应用。

参考资料

在 2.1 节中有关人眼结构的阅读资料可在 Atchison, Smith[2000]和 Oyster[1999]中找到。关于视觉感知的阅读资料见 Regan[2000]和 Gordon[1997]。Hubel[1988]和 Cornsweet[1970]的经典书籍也很重要。Born 和 Wolf[1999]是光和电磁波理论的基本参考资料。电磁能传播包含在 Felsen 和 Marcuvitz[1994]的论述中。

图像感知领域非常宽并且发展非常快。光和其他成像传感器的资料主要来自国际光学工程学会(SPIE)。下边是 SPIE 在这一领域有代表性的出版物: Blouke 等[2001], Hoover 和 Doty [1996]和 Freeman[1987]。

2.3.4 节提供的图像模型来自 Oppenheim, Schafer 和 Stockham[1968]的论著。本书所用的照度、反射值的参考文献是 IES 光学手册[2000]。图像取样和某些影响(如混淆)方面的其他读物见 Bracewell[1995]。2.4.3 节提到的感知图像质量(作为取样和量化的函数)方面的早期实验是由 Huang[1965]报告的。减少图像取样数和灰度级的同时尽量减小随之而来的退化方面的出版物一直是当前人们感兴趣的内容,如 Papamarkos 和 Atsalakis[2000]的例证。图像放大和收缩方面更进一步的读物见 Sid-Ahmed[1995], Unser 等[1995], Umbaugh[1998]和 Lehmann 等[1999]的文献。2.5 节所研究问题的读物见 Rosenfeld 和 Kak[1982], Marchand-Maillet 和 Sharaiha [2000], Ritter 和 Wilson[2001]提供的材料。图像处理中线性系统的读物可在 Castleman[1996]中找到。

习题

★2.1 使用 2.1 节提供的背景信息,并采用纯几何方法,如果纸上的打印点离眼睛 0.2 m 远,估计眼睛能辨别的最小打印点的直径。为了简单起见,假定当在黄斑处的像点变得远比视网膜区域的接收器(锥状体)直径小的时候,视觉系统已不能检测到该点。进一步假定黄斑可用 $1.5\ \text{mm} \times 1.5\ \text{mm}$ 的方阵模型化,并且锥状体和锥状体间的空间在该阵列上均匀分布。

2.2 当在白天进入一个黑暗剧场时,在能看清并找到空座位时要用一段时间适应,2.1 节描述的视觉过程在这种情况下起什么作用?

★2.3 图 2.10 中没有显示交流电是电磁波谱的一部分,在美国,商用交流电的频率是 60 Hz。

这一波谱分量的波长是多少千米?

2.4 请设计一个研究细胞、细菌、病毒和蛋白质边界形状的图像系统前端。在这种情况下,前端由光源和相应的图像摄像机组成,要求每一类样品闭合环形的直径分别是 50,1,0.1 和 0.01 μm。

(a)能用单传感器和摄像机解决这一成像问题吗?

如果回答是肯定的,请详细说明需要的照明波长和摄像机?指出是彩色摄像机还是远红外摄像机,无论哪一种都给出合适的照明光源。

(b)如果在(a)中问题的回答是否定的,建议用哪种照明光源和相应的图像传感器?指出(a)所要求的光源和摄像机。需要用最小数量的照明光源和摄像机去解决这个问题。

2.5 7 mm×7 mm 的 CCD 芯片有 1 024×1 024 元素,将其聚焦到相距 0.5 m 远的方形平坦区域。该摄像机每毫米能解析多少线对?摄像机配置 35 mm 镜头。

(提示:成像处理模型如图 2.3 所示,摄像机镜头聚焦长度代替眼睛的聚焦长度。)

★2.6 汽车制造商在跑车限版线的保险杠上自动配置一定的标记成分。这一成分就是协调的颜色,因此,机器人为了选择合适的保险杠部件需要知道每辆车的颜色。模型仅用 4 种颜色:蓝、绿、红和白。请提出一个基于成像的解决方案。怎样解决自动决定每辆车颜色的问题?请记住,在选择部件时价格是最主要的考虑因素。

2.7 假设中心在(x_0,y_0)的平坦区域被一个强度分布为:

$$i(x,y) = Ke^{-[(x-x_0)^2+(y-y_0)^2]}$$

的光源照射。为简单起见,假设区域的反射是恒定的,并等于 1.0,令 $K=255$。如果图像用 k 比特的强度分辨率进行数字化,并且眼睛可检测相邻像素间 8 种灰度的突变,那么 k 取什么值将导致可见的伪轮廓?

2.8 对习题 2.7,在 $k=2$ 时画一个草图。

★2.9 数字数据传输通常用波特率度量,其定义为每秒钟传输的比特数。通常的传输是以一个开始比特、一个字节(8 比特)的信息和一个停止比特组成的包完成的。基于这个概念回答下列问题:

(a)用 56 K 波特的调制解调器传输一幅 1 024×1 024、256 级灰度的图像要花费几分钟?

(b)以 750 K 波特[是典型的电话 DSL(数字用户线)连接的速度]传输要用多少时间?

2.10 高清晰度电视(HDTV)用 1 125水平电视线的分辨率隔行扫描产生图像(每隔一行在显像管表面画一条线,每两场形成一帧,每场用一秒钟的1/60时间)。图像宽高的横纵比是 16:9。用清楚的水平线确定图像的垂直分辨率。一个公司已经设计了一种图像获取系统,该系统从 HDTV 图像产生数字图像。在该系统中每条 TV(水平)线的分辨率与垂直分辨率成比例,该比例是图像的宽高比。彩色图像的每个像素都有 24 比特的强度分辨率,红、绿、蓝图像的每个像素都有 8 比特。这三幅原色图像形成彩色图像。存储 2 小时的 HDTV 节目将使用多少比特?

★2.11 两个图像子集 S_1 和 S_2 如下图所示。对于 $V=\{1\}$,确定这两个子集是(a)4 邻接,(b)8 邻接,还是(c)m 邻接。

	S_1				S_2				
0	0	0	0	0	0	0	1	1	0
1	0	0	1	0	0	1	0	0	1
1	0	0	1	0	1	1	0	0	0
0	0	1	1	1	0	0	0	0	0
0	0	1	1	1	0	0	1	1	1

★2.12 对将一个像素宽度的 8 通路转换到 4 通路提出一种算法。

2.13 对将一个像素宽度的 m 通路转换到 4 通路提出一种算法。

2.14 说明如 2.5.2 节定义的区域边界是闭合通路。

★2.15 考虑如下所示的图像分割：

(**a**)令 $V=\{0,1\}$ 并计算 p 和 q 间的 4,8,m 通路的最短长度。如果在这两点间不存在特殊通路，请解释原因。

(**b**)对 $V=\{1,2\}$ 重复上题。

$$
\begin{array}{ccccc}
 & 3 & 1 & 2 & 1\ (q) \\
 & 2 & 2 & 0 & 2 \\
 & 1 & 2 & 1 & 1 \\
(p) & 1 & 0 & 1 & 2
\end{array}
$$

★2.16 (**a**)对于 p 和 q 两点间的 D_4 距离等于这两点间最短 4 通路的情况，给出需要的条件。

(**b**)这个通路惟一吗?

2.17 对 D_8 距离重复习题 2.16 的问题。

★2.18 在下一章中我们将讨论算子，其函数在一个很小的子图像区 S 计算像素总数。说明这些都是线性算子。

2.19 一个数集的中值 ξ 表示这样一个数：数集的一半数值比它大，另一半比它小。例如，数集{2,3,8,20,21,25,31}的中值是 20。说明计算子图像区域 S 中值的算子是非线性的。

2.20 一个工厂生产半透明的微型聚合物方形线。要求进行 100%视觉检测来严格控制质量，工厂管理者发现用人检测越来越贵，于是检测改为半自动的。在每一个检测位置，机械装置把聚合方形物放在一个灯上方，灯位于光学系统的下方，该光学系统产生方形物的放大图像，图像完全落在 80 mm×80 mm 的观察屏上。缺陷以较暗的小圆点形式出现，检测者的工作是观察屏幕，如果一个样品在屏幕的测量尺度上，有多个直径大于或等于 0.8 mm 的这种小圆点，则要将该样品剔除。管理者发现，如果能找到一个全自动的处理方法，利润将提高 50%。她还相信如果项目成功，这将是帮助她升为法人的阶梯。经过大量调研之后，管理者决定了解决这个问题的方法，即用 CCD TV 摄像机去观察每个检测屏幕，并把摄像机的输出送人能够检测小圆点的图像处理系统，测量它们的直径，并触发以前由检测员操作的接收和剔除按钮。并且可以找到完成这种工作的系统，只要在数字图像中最小缺陷至少占有 2×2 像素面积。管理者聘任你帮助她拟定摄像机和镜头系统的技术条件，但是要求使用现有的元件。对于镜头来说，意味着任何 25 mm，35 mm 直至 200 mm 的整数倍的焦距。对于摄像机，其意思是分辨率为 512×512，1 024×1 024 或 2 048×2 048 像素。这些摄像机单个成像元素的方形

尺寸为8 μm×8 μm,成像元素间的距离是2 μm。对于这些条件,摄像机远比镜头要贵,因此,应该以镜头的选择为基础,用分辨率尽可能低的摄像机去解决这个问题。你可以对这个项目提出一些建议,详细叙述你的分析并得出结论。使用习题2.5提出的成像几何结论。

第3章　空间域图像增强

增强的首要目标是处理图像，使其比原始图像更适合于特定应用。这里的"特定"很重要，因为它一开始就确立了本章所讨论技术是面向问题的。例如，一种很适合增强X射线图像的方法，不一定是增强由空间探测器发回的火星图像的最好方法。暂且不谈所用的方法，图像增强本身就是图像处理中最具有吸引力的领域之一。

图像增强的方法分为两大类：空间域方法和频域方法。"空间域"一词是指图像平面自身，这类方法是以对图像的像素直接处理为基础的。"频域"处理技术是以修改图像的傅氏变换为基础的。空间域方法在这一章讲述，频域增强将在第4章讨论。以这两类方法的各种结合为基础的增强技术是不常见的。我们也注意到，本章关于增强的许多基本技术在后续章节里的其他图像处理应用中也会用到。

图像增强的通用理论是不存在的。当图像为视觉解释而进行处理时，由观察者最后判断特定方法的效果。图像质量的视觉评价是一种高度主观的过程，因此，定义一个"理想图像"标准，通过这个标准去比较算法的性能。当为机器感知而处理图像时，这个评价任务就会容易一些。例如，在一个特征识别的应用中，不考虑像计算要求这些问题，最好的图像处理方法是一种能得到最好的机器可识别结果的方法。不论怎样，甚至在把一个明确的性能标准加于这个问题的情况下，在选择特定的图像增强方法之前，常常需要一个试验和误差的特定量。

3.1　背景知识

如前所述，"空间域增强"是指增强构成图像的像素。空间域方法是直接对这些像素操作的过程。空间域处理可由下式定义：

$$g(x,y) = T[f(x,y)] \quad (3.1.1)$$

其中 $f(x,y)$ 是输入图像，$g(x,y)$ 是处理后的图像，T 是对 f 的一种操作，其定义在 (x,y) 的邻域。另外，T 能对输入图像集进行操作，例如，为减少噪音而对 K 幅图像进行逐像素的求和操作，如3.4.2节所讨论的。

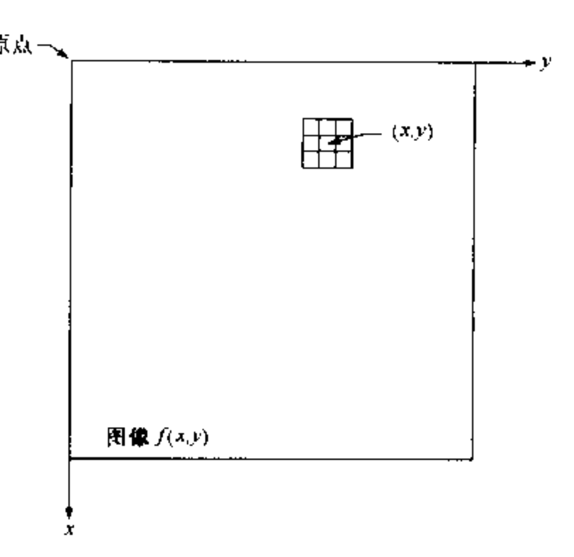


图3.1　图像中 (x,y) 点的 3×3 邻域

定义一个点 (x,y) 邻域的主要方法是利用中心在 (x,y) 点的正方形或矩形子图像，如图3.1所示。子图像的中心从一个像素向另一个像素移动，比如说，可以从左上角开始。T 操作应用到每一个 (x,y) 位置得到该点的输出 g。这个过程仅仅用在小范围邻域里的图像像素。尽管像近似于圆的其他邻域形状有时也用，但正方形和矩形阵列因其容易执行操作而占主导地位。

T 操作最简单的形式是邻域为 1×1 的尺度(即单个像素)。在这种情况下，g 仅仅依赖于

f 在(x,y)点的值，T 操作成为灰度级变换函数(也叫做强度映射)，形式为：

$$s = T(r) \tag{3.1.2}$$

这里，为简便起见，令 r 和 s 是所定义的变量，分别是 $f(x,y)$和 $g(x,y)$在任意点(x,y)的灰度级。例如，如果 $T(r)$有如图 3.2(a)所示的形状，这种变换将会产生比原始图像更高的对比度，进行变换时，在原始图像中，灰度级低于 m 时变暗，而灰度级在 m 以上时变亮。在这种对比度扩展技术里，在 m 以下的 r 值将被变换函数压缩在 s 的较窄范围内，接近黑色。对 m 以上的 r 值执行相反的操作。在极限情况下，如图 3.2(b)所示，$T(r)$产生了两级(二值)图像。这种形式的映射关系叫做阈值函数。有的相当简单，却有很大作用，处理方法可以用灰度变换加以公式化。因为在图像任意点的增强仅仅依赖于该点的灰度，这类技术常常是指点处理。

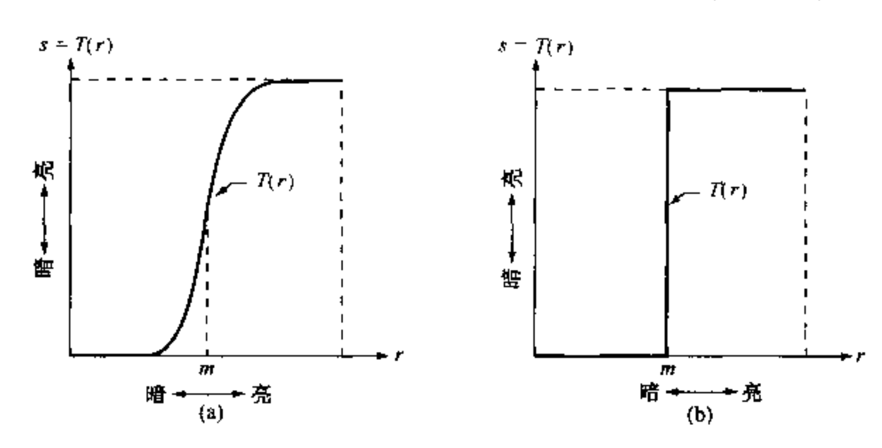


图 3.2 对比度增强的灰度级变换函数

更大的邻域会有更多的灵活性。一般的方法是，利用点(x,y)事先定义的邻域里的一个 f 值的函数来决定 g 在(x,y)的值，其公式化的一个主要方法是以利用所谓的模板(也指滤波器、核、掩模或窗口)为基础的。从根本上说，模板是一个小的(即 3×3)二维阵列，如图 3.1 所示，图中，模板的系数值决定了处理的性质，如图像尖锐化等。以这种方法为基础的增强技术通常是指模板处理或滤波。这些概念将在 3.5 节讨论。

3.2 某些基本灰度变换

以讨论灰度变换函数开始研究图像增强技术，这些都属于所有图像增强技术中最简单的一类。处理前后的像素值用 r 和 s 分别定义。如前节所述，这些值与 $s = T(r)$表达式的形式有关，这里的 T 是把像素值 r 映射到值 s 的一种变换。由于处理的是数字量，变换函数的值通常储存在一个一维阵列中，并且从 r 到 s 的映射通过查表得到。对于 8 比特环境，一个包含 T 值的可查阅的表需要有 256 个记录。

正如对灰度变换介绍的那样，考虑图 3.3，它显示了图像增强常用的三个基本类型函数：线性的(正比和反比)、对数的(对数和反对数变换)、幂次的(n 次幂和 n 次方根变换)。正比函数是最一般的，其输出亮度与输入亮度可互换，惟有它完全包括在图形中。

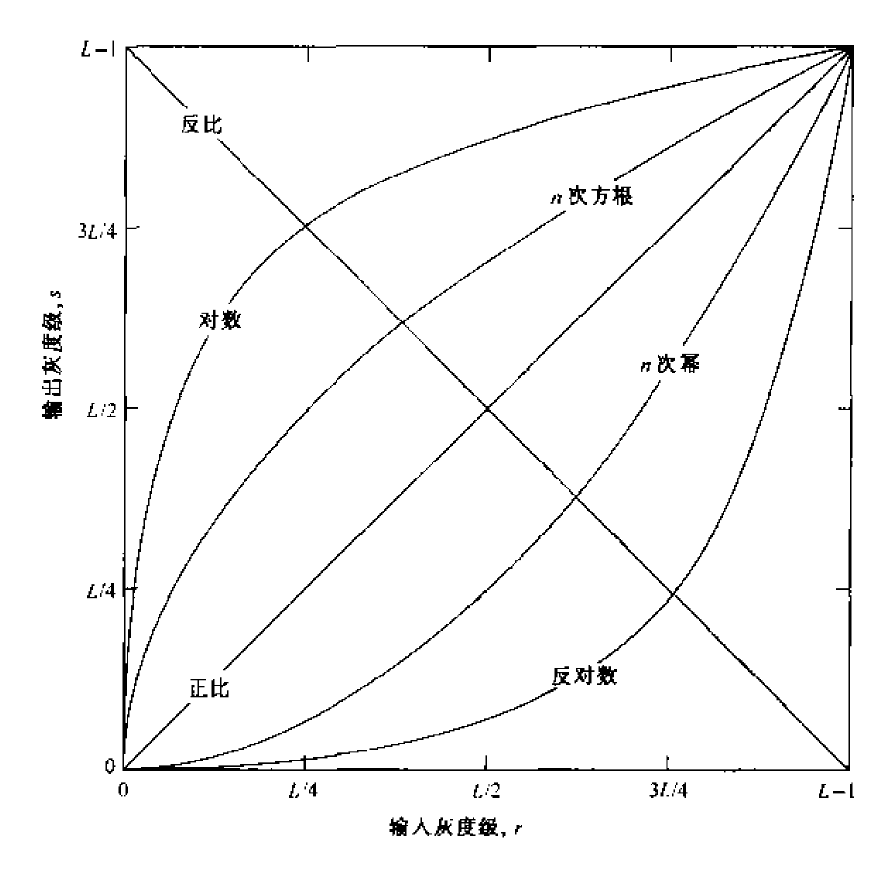


图3.3　用于图像增强的某些基本灰度变换函数

3.2.1　图像反转

灰度级范围为[0, L−1]的图像反转可由示于图3.3的反比变换获得，表达式为：

$$s = L - 1 - r \tag{3.2.1}$$

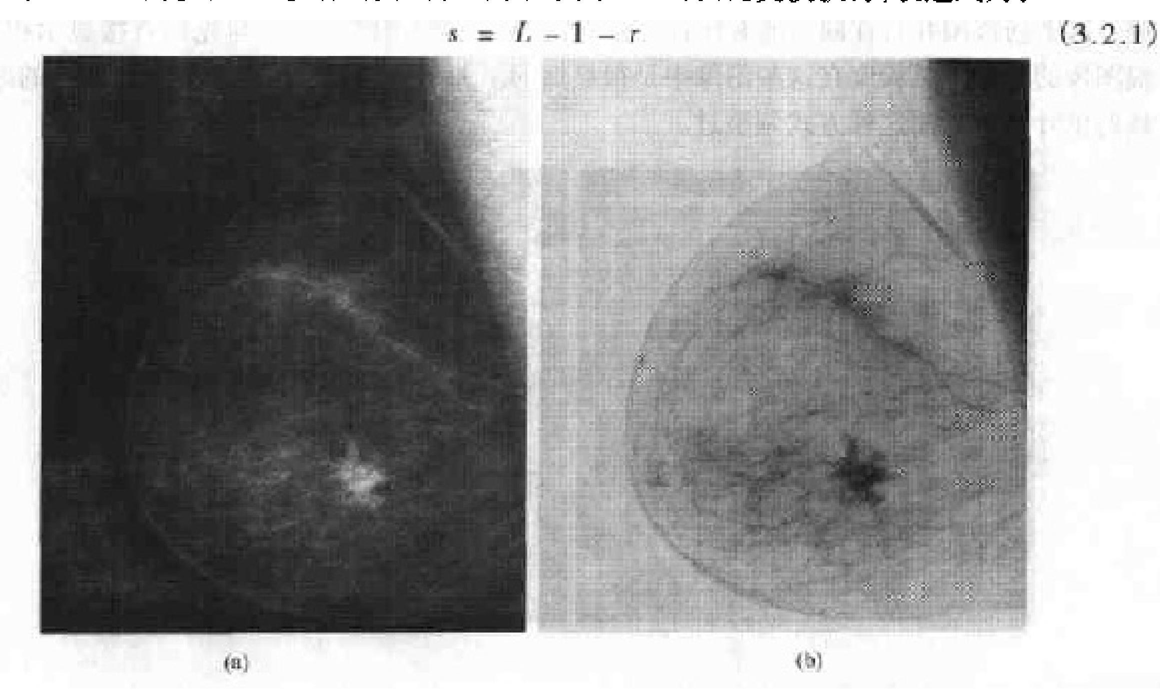


图3.4　(a)原始数字乳房照片，(b)用式(3.2.1)反变换得到的反转图像(GE医学系统提供)

用这种方式倒转图像的强度产生图像反转的对等图像。这种处理尤其适用于增强嵌入于图像暗色区域的白色或灰色细节，特别是当黑色面积占主导地位时。一个例子示于图 3.4，原始图像为一乳房的数字 X 照片，可看到有一小块病变。尽管事实上两幅图在视觉内容上都一样，但注意，在这种特殊情况下，分析乳房组织结构时反转图像要容易得多。

3.2.2 对数变换

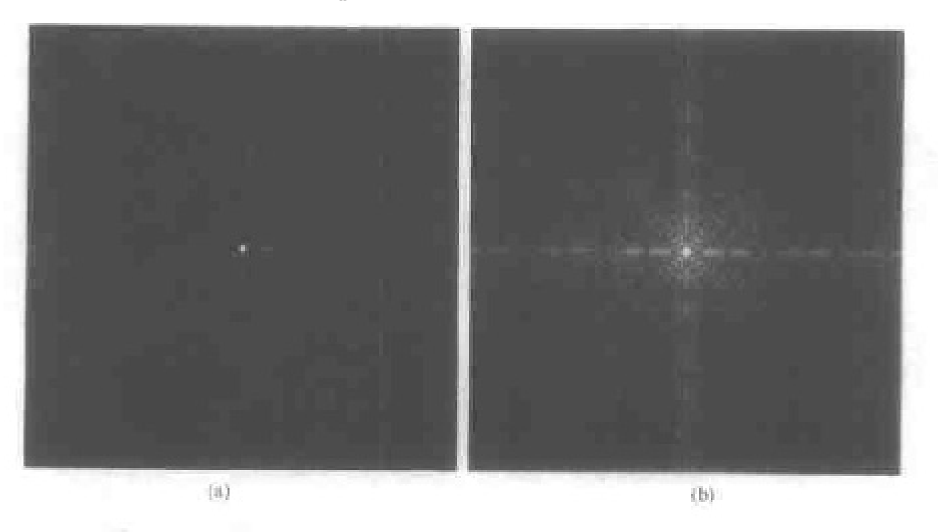
示于图 3.3 的对数变换的一般表达式为：

$$s = c\log(1 + r) \tag{3.2.2}$$

其中 c 是一个常数，并假设 $r \geqslant 0$。对数曲线如图 3.3 所示。此种变换使一窄带低灰度输入图像值映射为一宽带输出值。相对的是输入灰度的高调整值。可以利用这种变换来扩展被压缩的高值图像中的暗像素。相对的是反对数变换的调整值。

如图 3.3 所示的一般对数函数的所有曲线都能完成图像灰度的扩散/压缩。事实上，就此目的而言，下节将讨论的幂次规则变换比对数变换更加灵活。不管怎样，对数函数有它重要的特征，就是它在很大程度上压缩了图像像素值的动态范围，其应用的一个典型例子就是傅里叶频谱，它的像素值有很大的动态范围，这将在第 4 章中讨论。现在，我们只注意图像频谱的特征。频谱值的范围从 0 到 10^6 或更高的情况是不常见的。当计算机处理像这样的无误数字时，图像显示系统通常不能如实地再现如此大范围的强度值。最后的效果是有很多的细节会在典型的傅里叶频谱显示时丢失。

为了说明对数变换，图 3.5(a)显示了值为 $0 \sim 1.5\times10^6$ 的傅里叶频谱。当这些值在一个 8 比特的系统中被线性标度而显示时，最亮的像素将成为显示的重点，频谱中的低值(恰恰是重要的)将损失掉。这种显示重点的效果在图 3.5(a)中相对小的图像范围里鲜明地体现出来，而作为黑色则观察不到。如果不用这种方法显示数值，可以先对光谱值利用式(3.3.2)处理(在这种情况下，$c = 1$)，那么结果值由 0 变为一个易处理的值 6.2。图 3.5(b)显示了线性地调节这个新范围并且在同样的 8 比特显示系统中显示频谱的结果。与光谱直接显示相比，这幅图像的细节可见程度在这些图像中是很显然的。关于图像处理的出版物中所看到的绝大多数傅里叶频谱都用这种方式调整过。



(a) (b)

图 3.5 (a)傅里叶频谱，(b)对式(3.2.2)取 $c = 1$，应用了对数变换的结果



3.2.3 幂次变换

幂次变换的基本形式为：

$$s = cr^{\gamma} \tag{3.2.3}$$

其中 c 和 γ 为正常数。有时考虑到偏移量（即当输入为 0 时的可测量输出），式(3.2.3)也写做 $s = c(r+\varepsilon)^{\gamma}$。不管怎样，偏移量通常是显示标定的衍生，并且一般在式(3.2.3)中忽略掉。作为 r 的函数，s 对于 γ 的各种值绘制的曲线示于图 3.6。如对数变换的情况一样，幂次曲线中 γ 的部分值把输入窄带暗值映射到宽带输出值。相反，输入高值时也成立。然而，不像对数函数，我们注意到这里随着 γ 值的变化将简单地得到一族变换曲线。如预期的一样，我们看到图 3.6 中 $\gamma > 1$ 的值和 $\gamma < 1$ 的值产生的曲线有相反的效果。最后，我们应注意式(3.2.3)当 $c = \gamma = 1$时，将简化为正比变换。

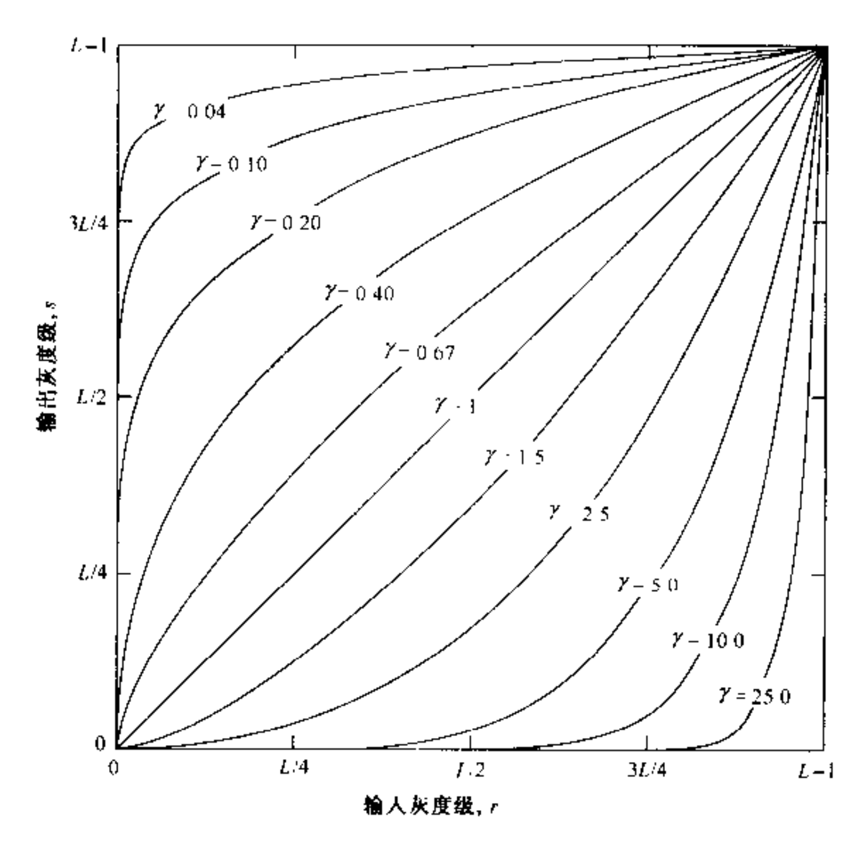


图 3.6　对于各种 γ 值(所有情况，$c = 1$)，式 $s = cr^{\gamma}$ 的曲线

用于图像获取、打印和显示的各种装置根据幂次规律进行响应。习惯上，幂次等式中的指数是指伽马值[因此在式(3.2.3)中用到这一符号]。用于修正幂次响应现象的过程称做伽马校正。例如，阴极射线管(CRT)装置有一个电压 – 强度响应，这是一个指数变化范围为 1.8～2.5 的幂函数。在图3.6 中，用 $\gamma = 2.5$ 的参考曲线，我们看到这样的显示系统倾向于产生比希望的效果更暗的图像。这个结果可由图 3.7 进行说明。图 3.7(a)显示一个简单的灰度线性楔形输入到 CRT 监视器。如预期的那样，CRT 显示器的输出比输入暗，如图 3.7(b)所示。在这种情况下伽马校正很简单，需要做的只是将图像输入到监视器前进行预处理，即进行 $s = r^{1/2.5} = r^{0.4}$ 的变换，其结果如图 3.7(c)所示。当输入同样的监视器时，这一伽马校正的输入将产生接

近于原图像的输出,正如图3.7(d)所示。类似的分析可用于其他图像装置,如扫描仪和打印机。惟一的不同就是随装置而定的伽马值(Poynton[1996])。

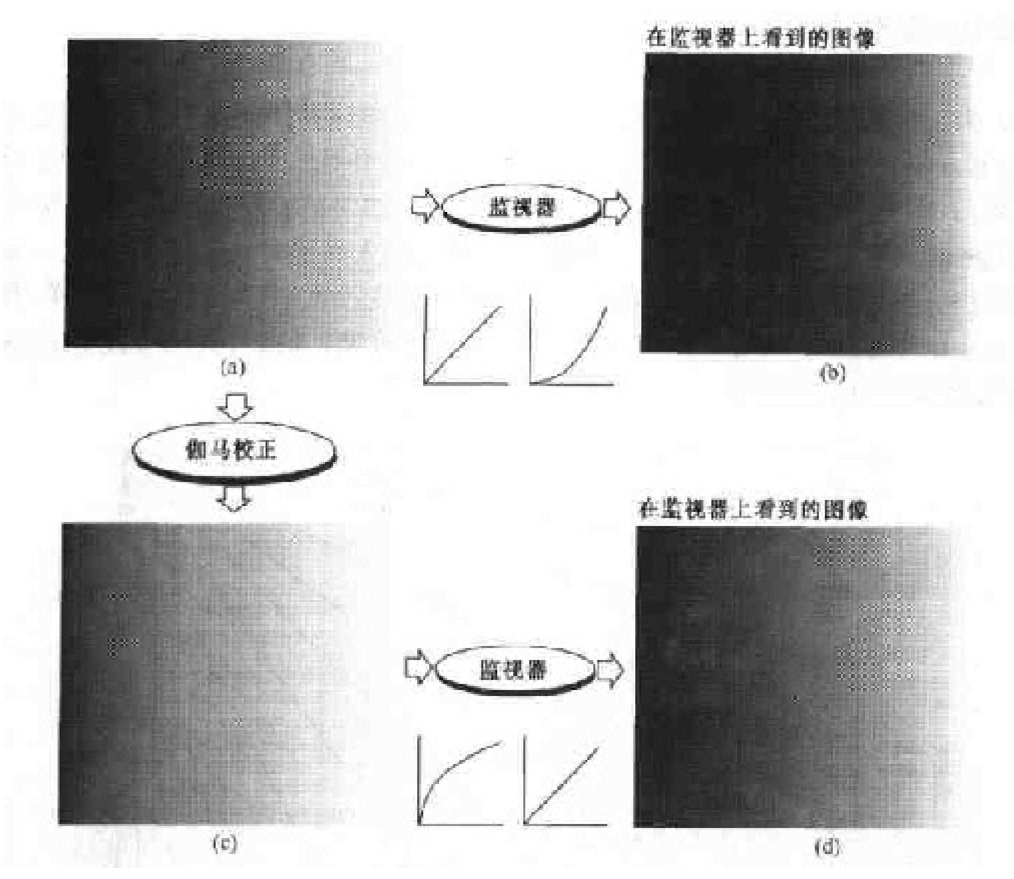


图3.7 (a)线性楔形灰度图像,(b)对线性楔形灰度图像的监视器响应,(c)伽马校正楔形图像,(d)监视器输出

如果涉及在计算机屏幕上精确显示图像,伽马校正是很重要的。不恰当的图像修正会被漂白或变得更暗。试图精确再现颜色也需要伽马校正的一些知识,这是因为改变伽马校正值不仅可改变亮度,还可改变红、绿、蓝的比率。随着数字图像在因特网上商业应用的增多,在过去几年里,伽马校正逐渐变得越来越重要。对于成百上千万的网民(这些人的绝大多数都有不同的监视器或监视器设置)浏览的流行网站,为其创作图像是经常的事。有些计算机系统甚至配有部分伽马校正。同时,目前的图像标准没有包括创作图像的伽马校正值,因此,问题更加复杂化了。由于这些限制,当在网站中存储图像时,一个可能的方法就是用伽马值对图像进行预处理,此伽马值表示了在开放的市场中,在任意给定时间点,各种型号的监视器和计算机系统所被期望的"平均值"。

例3.1 用幂次变换进行对比度增强

除了伽马校正之外,幂次变换在一般用途的对比度操作中也很有用。图3.8(a)中显示了一幅人体胸上部脊椎骨折和椎线受影响的核磁共振(MR)图像。在胸椎垂直中心附近,即图中上部约1/4处,骨折显而易见,由于所给图像整体为暗色,灰度的扩大是需要的,这可由指数为分数的幂次变换来完成。对图3.8(a)使用式(3.2.3)的幂次变换函数处理可以得到示于图中的其他几幅图像。图3.8(b)到(d)相应的伽马值分别为0.6,0.4

和 0.3(c 始终为 1)。注意到伽马值从 0.6 到 0.4 时，可看见更多的细节，当伽马值进一步减少到 0.3 时，背景中细节进一步增强，但图像看上去开始有轻微的“冲淡”外观的那一点的对比度开始减小，尤其是背景。比较所有结果，我们看到对比度和可辨别细节的最好增强效果在 $\gamma=0.4$时。在这幅特定的图像中，当略低于 $\gamma=0.3$ 时，对比度会减到难以接受的程度。

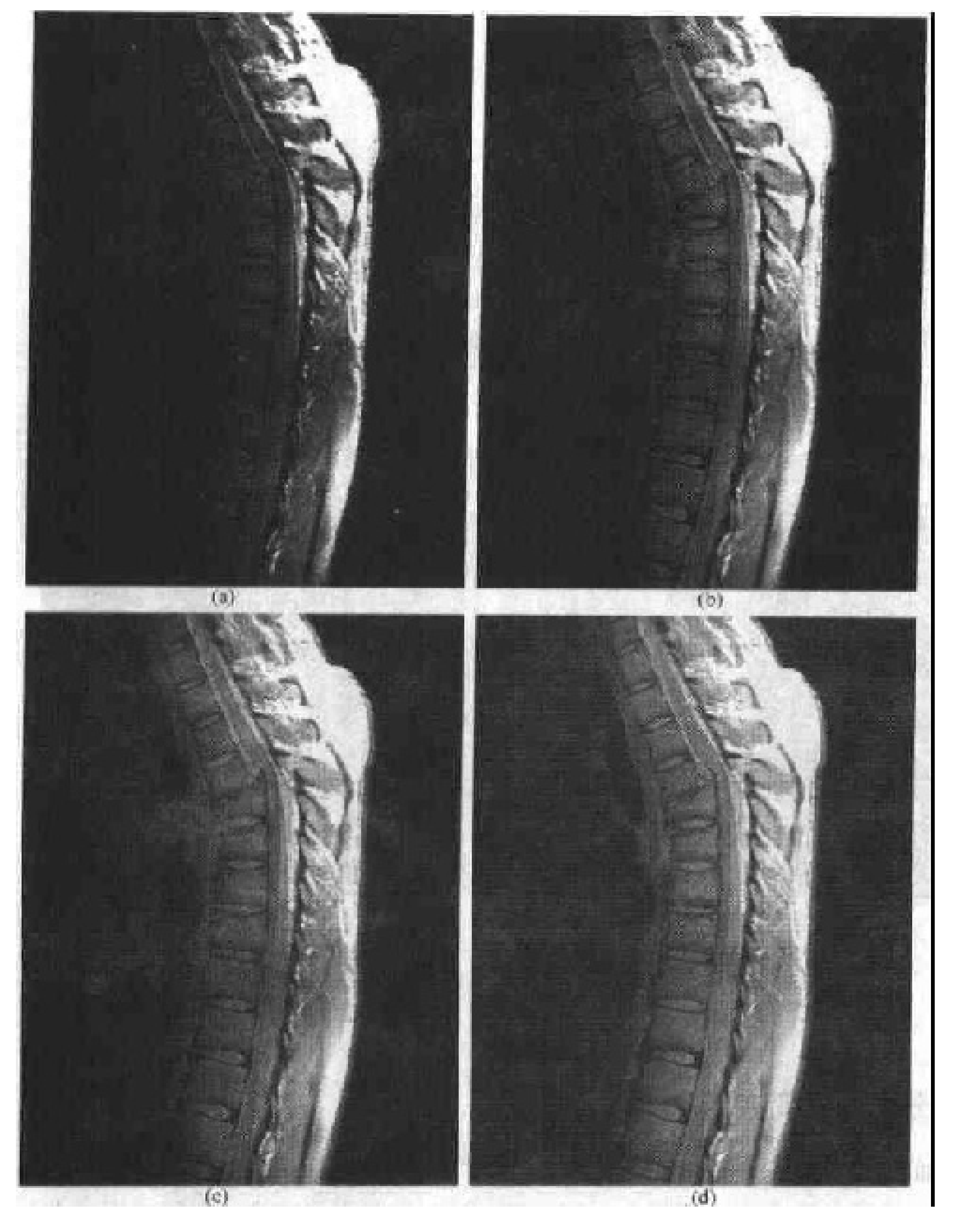


图 3.8　(a)人的脊椎骨折的核磁共振图像，(b)～(d)应用式(3.2.3)并且 $c=1,\gamma$ 分别等于0.6、0.4，0.3时变换的结果(这个例子的原图像由Vanderbilt大学医学中心放射学和辐射学系的David R. Pickens博士提供)

例 3.2　幂次变换的另一例证

图 3.9(a)是图 3.8(a)一个相对的问题。现在将要增强的图像有“冲淡”的显示效果，表明

灰度级的压缩是必需的。这可以用式(3.2.3),并且令 $c=1,\gamma$ 值大于1来完成。令 $\gamma=3.0$,4.0和5.0的处理结果示于图3.9(b)到图3.9(d)。伽马值取3.0和4.0时,可得到合适的结果,且后者由于有较高的对比度而显示出较好的效果。$\gamma=5.0$ 得到的结果有些地方太暗,从而丢失了一些细节,如左上方主道路左部的暗区就是这样的例子。

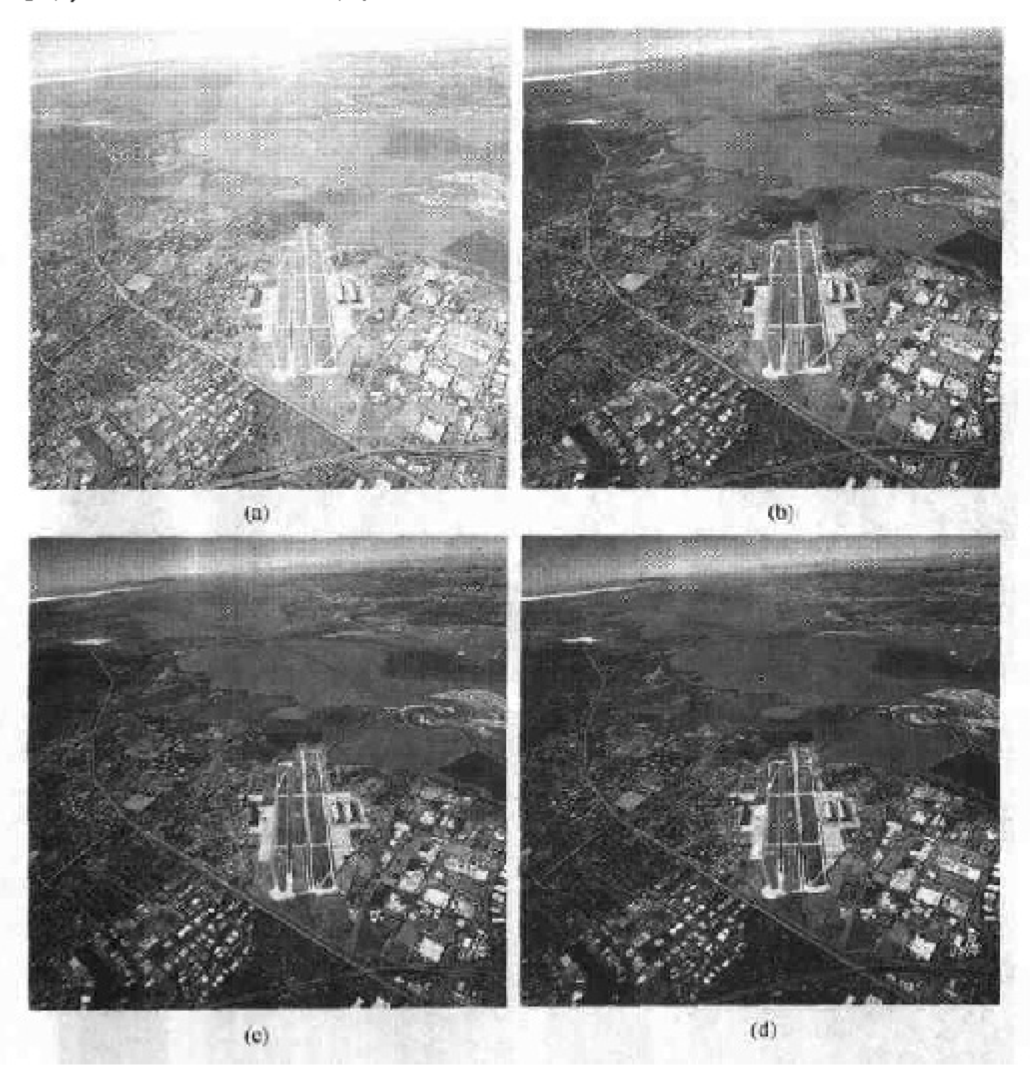


图3.9　(a)航空图像,(b)~(d)用式(3.2.3)并且取 $c=1,\gamma=3.0$,4.0和5.0变换的结果(此例的原图像由NASA提供)

3.2.4　分段线性变换函数

对前面三小节中所讨论方法的补充是分段线性函数。其相比前面所讨论函数的主要优势在于它的形式可任意合成。事实上,可以立刻看到,有些重要变换的实际应用可由分段线性函数描述。分段线性函数的主要缺点是其需要更多的用户输入。

对比拉伸

最简单的分段线性函数之一是对比拉伸变换。低对比度图像可由照明不足、成像传感器动态范围太小,甚至在图像获取过程中透镜光圈设置错误引起。对比拉伸的思想是提高图像

处理时灰度级的动态范围。

图3.10(a)是对比拉伸的典型变换。点(r_1, s_1)和(r_2, s_2)的位置控制了变换函数的形状。如果$r_1 = s_1$且$r_2 = s_2$，变换为一线性函数，它产生一个没有变化的灰度级。若$r_1 = r_2$，$s_1 = 0$，且$s_2 = L-1$，变换变为阈值函数，并产生二值图像，如图3.2(b)所示。(r_1, s_1)和(r_2, s_2)的中间值将产生输出图像中灰度级不同程度的展开，因而影响其对比度。一般情况下，假定$r_1 \leqslant r_2$且$s_1 \leqslant s_2$，函数则为单值单调增加。这样将保持灰度级的次序，因此避免了在处理过的图像中产生人为强度。

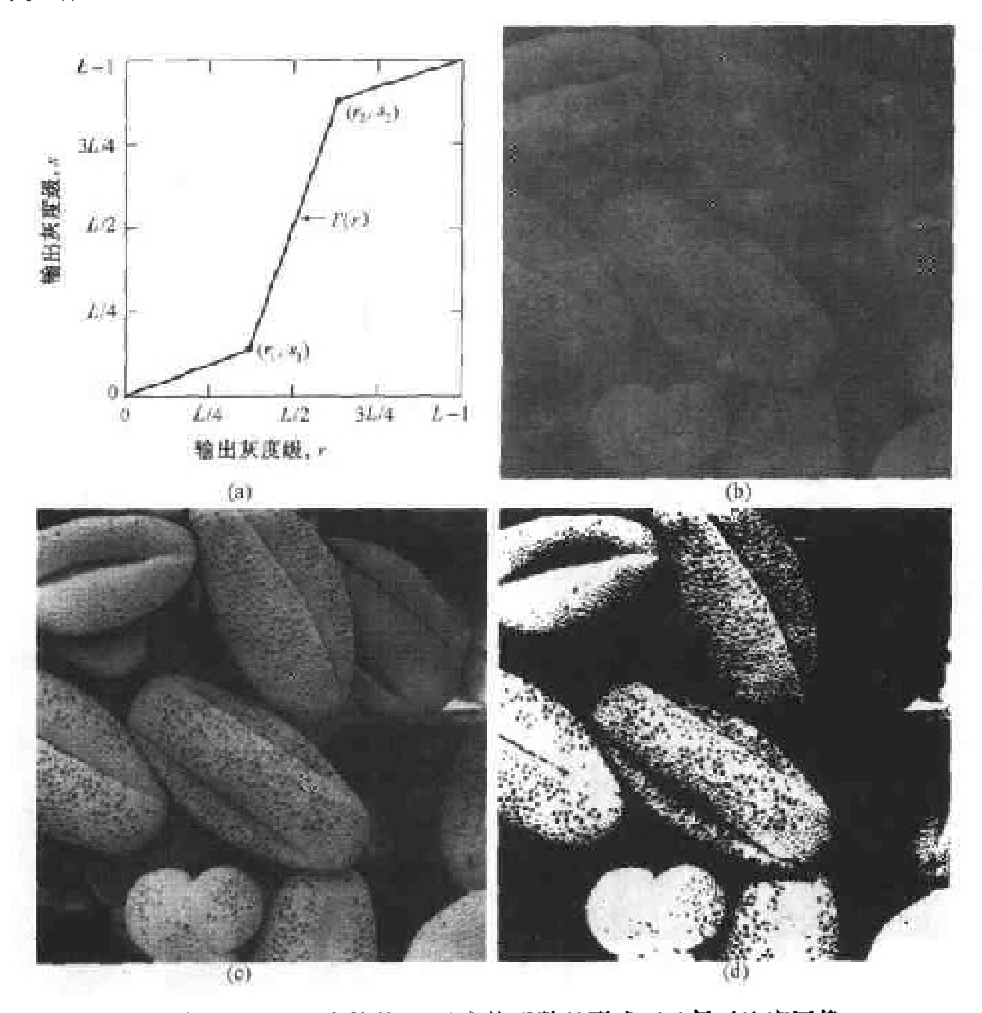


图3.10　对比度拉伸。(a)变换函数的形式，(b)低对比度图像，(c)对比度拉伸的结果，(d)门限化的结果(原图像由澳大利亚国立大学生物科学研究院Roger Heady博士提供)

图3.10(b)为一幅8比特低对比度图像。图3.10(c)为对比拉伸后的效果，设置$(r_1, s_1) = (r_{min}, 0)$且$(r_2, s_2) = (r_{max}, L-1)$，其中$r_{min}$，$r_{max}$分别代表图像中灰度级的最小值和最大值。因此，变换函数把灰度级由原范围线性地拉伸至饱和范围$[0, L-1]$。最后，图3.10(d)显示了使用这种前面定义的阈值函数$[r_1 = r_2 = m$(即图像的平均灰度级)$]$的效果。原始图像为电子显微镜扫描的放大约700倍的花粉图像。

灰度切割

在图像中提高特定灰度范围的亮度通常是必要的,其应用包括增强特征(如卫星图像中大量的水)和增强 X 射线图中的缺陷。有许多方法可以进行灰度切割,但是,它们中的大多数是两种基本方法的变形。其一就是在所关心的范围内为所有灰度指定一个较高值,而为其他灰度指定一个较低值。如图 3.11(a)所示,这个变换产生了一个二进制图像。基于如图3.11(b)所示变换的第二种方法使所需范围的灰度变亮,但是仍保持了图像的背景和灰度色调。图 3.11(c)表示一个灰度图像,图 3.11(d)表示使用了图 3.11(a)的变换后的结果。图 3.11 显示的两种变换的不同很容易阐明。

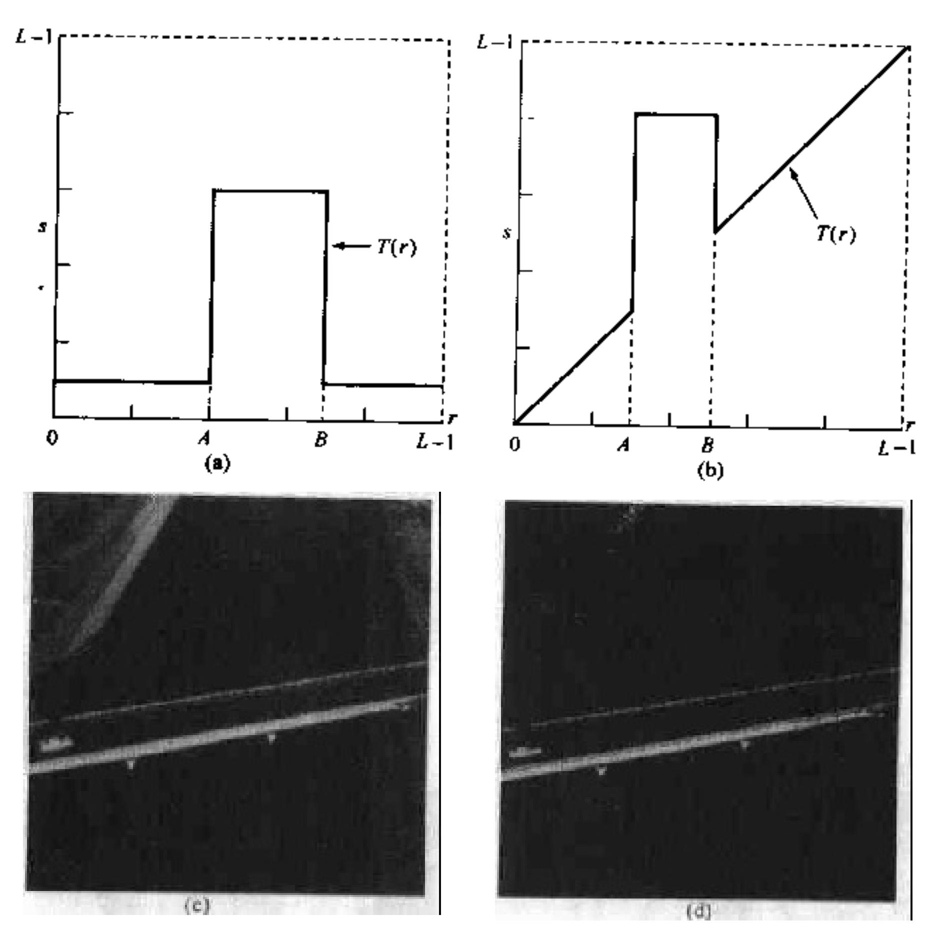


图 3.11 (a)这一变换加亮了 [A,B]范围的灰度级,所有其他灰度减小为一个恒定灰度级,(b)这一变换加亮了[A,B]范围的灰度,但保持所有其他灰度级不变,(c)一幅图像,(d)使用(a)变换的结果

位图切割

代替提高灰度范围的亮度,而通过对特定位提高亮度,对整幅图像质量仍然是有贡献的。设图像中的每一个像素都由 8 比特表示,假设图像是由 8 个 1 比特平面组成,其范围从最低有效位的位平面 0 到最高有效位的位平面 7。在 8 比特字节中,平面 0 包含图像中像素的最低位,而平面 7 则包含最高位。图 3.12 说明了这些概念,图 3.14 显示了在图 3.13 中描述的各种

位平面。注意到较高阶位(特别是前四位)包含了大多数在视觉上很重要的数据。其他位平面对图像中更多的微小细节有作用。把数字图像分解成为位平面,对于分析每一位在图像中的相对重要性是有用的,这是一个用来辅助决定量化一个像素的位数是否充足的过程。所以,这种分解在第 8 章所讨论的图像压缩中很有用。

就 8 比特图像的位平面抽取而言,说明用一个灰度阈值变换函数处理输入图像可以获得位平面 7 的二值图像并不困难。该灰度阈值变换函数:(1)把图像中 0 和 127 间的所有灰度映射到一个灰度级(例如 0);(2)把 129 到 255 间的灰度映射为另一种灰度级(例如 255)。在图 3.14中以位平面 7 表示的二值图像就是通过这种方式获得的。作为一个练习留给读者(习题 3.3),请算出能够产生其他位面的灰度变换函数。

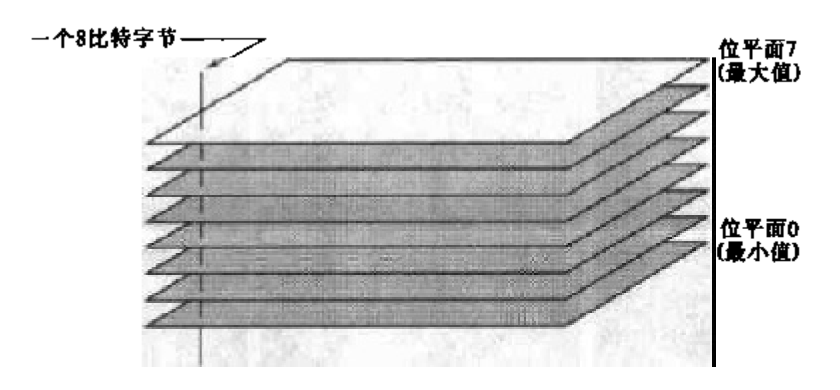


图 3.12　8 比特图像的位平面表示

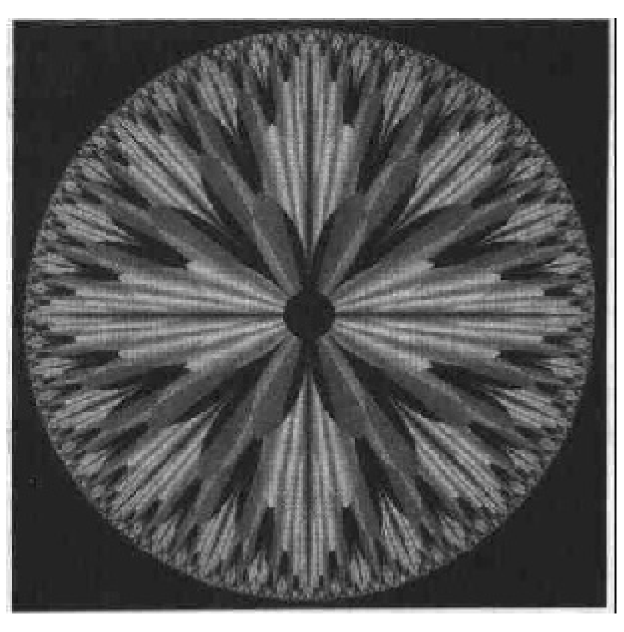


图 3.13　一幅 8 比特分形图像(分形是数学表达式产生的图像)[Swarthmore 学院(Swarthmore, PA)的 Melissa D. Binde 先生提供]

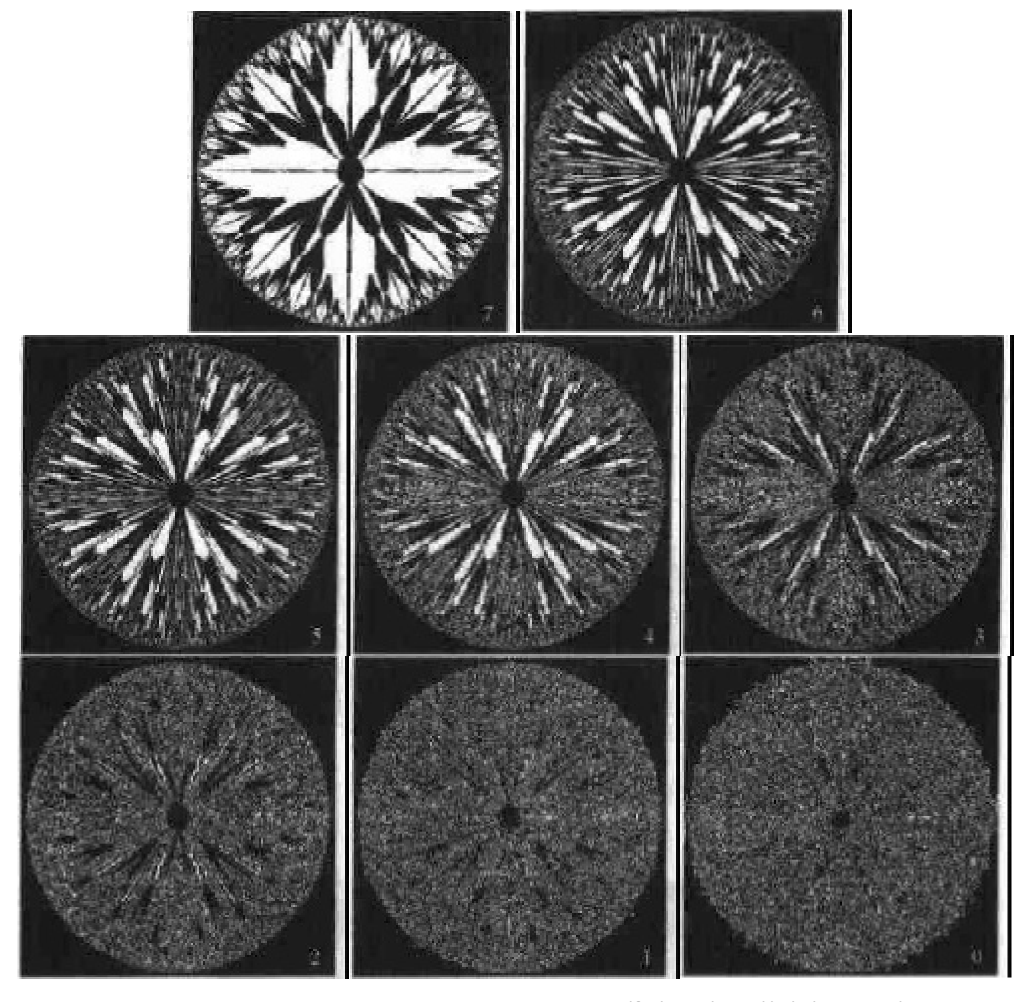


图 3.14 图 3.13 中图像的 8 个位平面。每一幅图像右下角的数字标识了位平面

3.3 直方图处理

灰度级为[0, $L-1$]范围的数字图像的直方图是离散函数 $h(r_k)=n_k$，这里 r_k 是第 k 级灰度，n_k 是图像中灰度级为r_k 的像素个数。经常以图像中像素的总数(用 n 表示)来除它的每一个值得到归一化的直方图。因此，一个归一化的直方图由 $P(r_k)=n_k/n$ 给出，这里 $k=0,1,\cdots,L-1$。简单地说，$P(r_k)$给出了灰度级为 r_k 发生的概率估计值。注意，一个归一化的直方图其所有部分之和应等于 1。

直方图是多种空间域处理技术的基础。直方图操作能有效地用于图像增强，如本节所示。除了提供有用的图像统计资料，在以后的章节会看到直方图固有的信息在其他图像处理应用中也是非常有用的，如图像压缩与分割。直方图在软件中易于计算，也适用于商用硬件设备，因此，它们成为了实时图像处理的一个流行工具。

作为在图像增强中直方图处理作用的一个介绍，考虑图 3.15，这是图 3.10 所示的以四个基本灰度级特征(暗、亮、低对比度和高对比度)做出的花粉图像。图的右侧显示了这些图像相应的直方图。每一个直方图曲线的水平轴对应灰度级值 r_k，纵轴对应于 $h(r_k)=n_k$的值或归

一化后为$p(r_k)=n_k/n$的值。如前边所指出的，这些直方图曲线仅仅是$h(r_k)=n_k$(作为r_k的函数)，或$P(r_k)=n_k/n$(作为r_k的函数)。

我们注意到在暗色图像中，直方图的组成成分集中在灰度级低(暗)的一侧。类似地，明亮图像的直方图则倾向于灰度级高的一侧。低对比度图像的直方图窄而集中于灰度级的中部。对于黑白图像，这意味着暗淡，好像灰度被冲淡了一样。最后，我们看到在高对比度的图像中，直方图的成分覆盖了灰度级很宽的范围，而且，像素的分布没有太不均匀，只有少量垂线比其他的高许多。直观上说，可以得出结论，若一幅图像其像素占有全部可能的灰度级并且分布均匀，则这样的图像有高对比度和多变的灰度色调。它的净作用是出现一幅灰度级丰富且动态范围大的图像。很快将显示出开发一种变换函数是可能的，该函数仅仅依靠输入图像直方图中的信息就可自动达到这一效果。

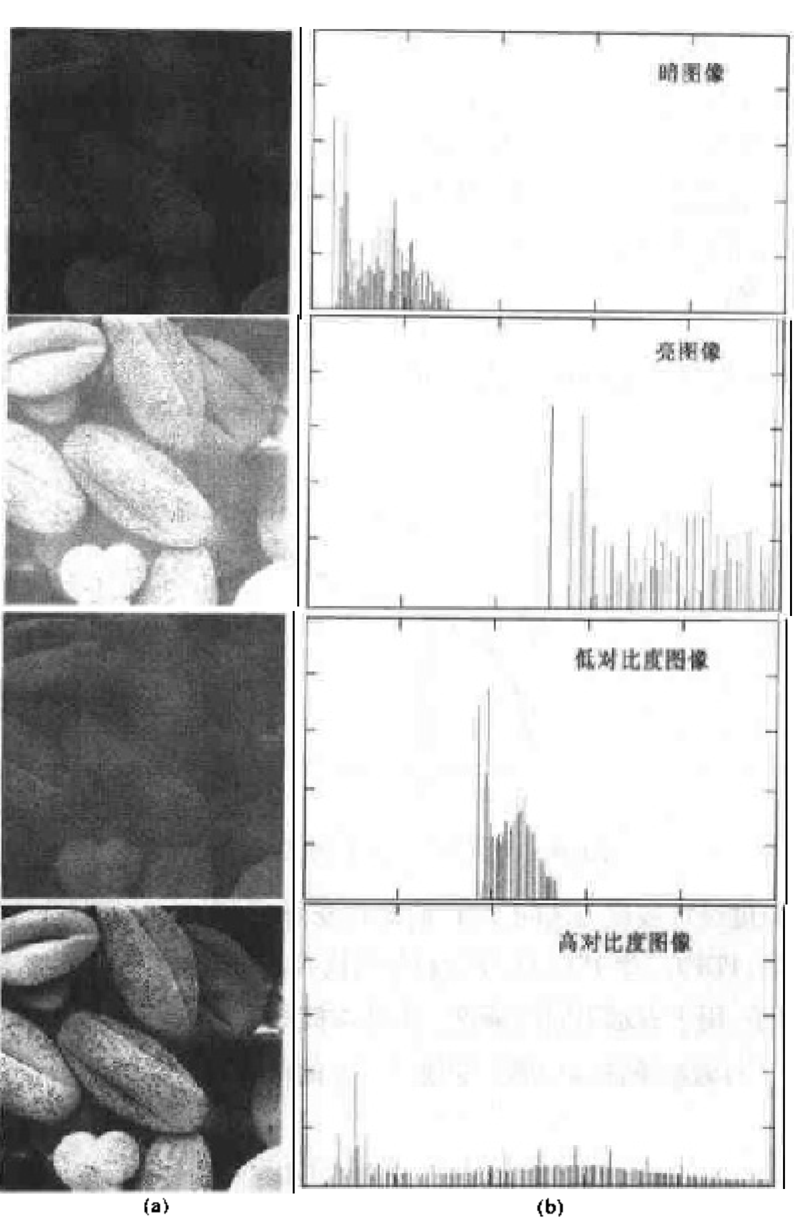


图3.15 四个基本图像类型，暗、亮、低对比度和高对比度以及与它们相对应的直方图（原图像由澳大利亚国立大学生物科学研究学院的Roger Heady博士提供）

3.3.1 直方图均衡化

考虑连续函数并且让变量 r 代表待增强图像的灰度级。在前面的讨论中,假设 r 被归一化到区间[0,1],且 $r=0$ 表示黑色及 $r=1$ 表示白色。然后,考虑一个离散公式并允许像素值在区间[0, L-1]内。

对于任一个满足上述条件的 r,我们将注意力集中在变换形式上:

$$s = T(r) \quad 0 \leqslant r \leqslant 1 \tag{3.3.1}$$

在原始图像中,对于每一个像素值 r 产生一个灰度值 s。显然,可以假设变换函数 $T(r)$ 满足以下条件:

(a) $T(r)$ 在区间 $0 \leqslant r \leqslant 1$ 中为单值且单调递增

(b) 当 $0 \leqslant r \leqslant 1$ 时, $0 \leqslant T(r) \leqslant 1$

条件(a)中要求 $T(r)$ 为单值是为了保证反变换存在,单调条件保持输出图像从黑到白顺序增加。变换函数不单调增加将导致至少有一部分亮度范围被颠倒,从而在输出图像中产生一些反转灰度级。这也许在某些时候需要使用,但在这里我们不需要。最后,条件(b)保证输出灰度级与输入有同样的范围。图3.16给出了满足这两个条件的一个变换函数的例子。由 s 到 r 的反变换可以表示为:

$$r = T^{-1}(s) \quad 0 \leqslant s \leqslant 1 \tag{3.3.2}$$

从例子(习题3.8)可得出,即使 $T(r)$ 满足条件(a)和(b),相应的函数 $T^{-1}(r)$ 也可能不为单值。

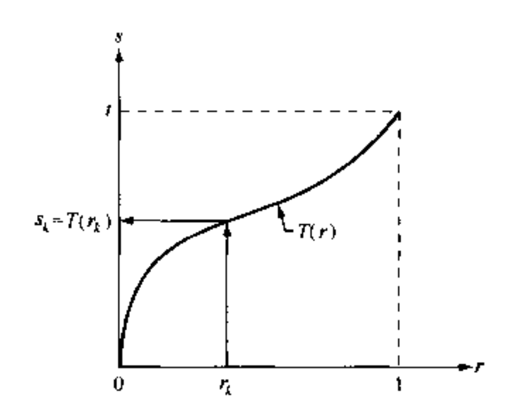


图3.16 单值单调递增的灰度级变换函数

一幅图像的灰度级可被视为区间[0,1]的随机变量。随机变量的一个最重要的基本描述是其概率密度函数(PDF)。令 $P_r(r)$ 和 $P_s(s)$ 分别代表随机变量 r 和 s 的概率密度函数。此处带有下标的 P_r 和 P_s 用于表示不同的函数。由基本概率理论得到一个基本结果:如果 $P_r(r)$ 和 $T(r)$ 已知,且 $T^{-1}(s)$ 满足条件(a),那么变换变量 s 的概率密度函数 $P_s(s)$ 可由以下简单公式得到:

$$P_s(s) = P_r(r)\left|\frac{\mathrm{d}r}{\mathrm{d}s}\right| \tag{3.3.3}$$

因此,变换变量 s 的概率密度函数由输入图像的灰度级PDF和所选择的变换函数决定。

在图像处理中一个尤为重要的变换函数如下所示:

$$s = T(r) = \int_0^r p_r(w)\mathrm{d}w \tag{3.3.4}$$

其中 w 是积分变量。式(3.3.4)的右部为随机变量 r 的累积分布函数(CDF)。因为概率密度函数永远为正,并且函数积分是一个函数曲线下的面积,所以它遵循该变换函数是单值单调增加的条件,因此,满足条件(a)。类似地,区间[0,1]上变量的概率密度函数的积分也在区间[0,1]上,因此,也满足条件(b)。

给定变换函数 $T(r)$,通过式(3.3.3)得到 $P_s(s)$。从基本微积分学(莱布尼茨准则),我们知道关于上限的定积分的导数就是该上限的积分值。也就是说:

$$\begin{aligned}\frac{\mathrm{d}s}{\mathrm{d}r} &= \frac{\mathrm{d}T(r)}{\mathrm{d}r} \\ &= \frac{\mathrm{d}}{\mathrm{d}r}\left[\int_0^r p_r(w)\mathrm{d}w\right] \\ &= p_r(r)\end{aligned} \tag{3.3.5}$$

用这个结果代替 $\mathrm{d}r/\mathrm{d}s$,代入式(3.3.3),取概率值为正,得到:

$$\begin{aligned}p_s(s) &= p_r(r)\left|\frac{\mathrm{d}r}{\mathrm{d}s}\right| \\ &= p_r(r)\left|\frac{1}{p_r(r)}\right| \\ &= 1 \qquad\qquad 0 \leqslant s \leqslant 1\end{aligned} \tag{3.3.6}$$

因为 $P_s(s)$是概率密度函数,在这里可以得出,区间[0,1]以外它的值为0,这是因为它在所有 s 值上的积分等于1。我们看到式(3.3.6)中给出的 $P_s(s)$形式为均匀概率密度函数。简而言之,已证明执行式(3.3.4)给出的变换函数会得到一随机变量 s,其特征为一均匀概率密度函数。特别要注意从式(3.3.4)得到 $T(r)$取决于 $P_r(r)$,但是,如式(3.3.6)指出的那样,$P_s(s)$的结果始终是均匀的,与 $P_r(r)$的形式无关。

对于离散值,我们处理其概率与和,而不是概率密度函数与积分。一幅图像中灰度级 r_k 出现的概率近似为:

$$p_r(r_k) = \frac{n_k}{n} \qquad k = 0,1,2,\cdots,L-1 \tag{3.3.7}$$

其中,如此节开始指出的,n 是图像中像素的总和,n_k 是灰度级为 r_k 的像素个数,L 为图像中可能的灰度级总数。式(3.3.4)中变换函数的离散形式为:

$$\begin{aligned}s_k = T(r_k) &= \sum_{j=0}^{k} p_r(r_j) \\ &= \sum_{j=0}^{k} \frac{n_j}{n} \qquad k = 0,1,2,\cdots,L-1\end{aligned} \tag{3.3.8}$$

因此,已处理的图像(即输出图像)由通过式(3.3.8),将输入图像中灰度级为 r_k 的各像素映射到输出图像中灰度级为 s_k 的对应像素得到。如前所述,作为 r_k的函数 $P_r(r_k)$的曲线称做直方图。式(3.3.8)给出的变换(映射)称做直方图均衡化或直方图线性化。不难得出式(3.3.8)的变换函数满足本节前边所述的条件(a)和(b)(见习题3.9)。

与连续形式不同,一般不能证明离散变换能产生均匀概率密度函数的离散值(为均匀直方图)。不论怎样,可以很容易地看出,式(3.3.8)的应用有展开输入图像直方图的一般趋势,以

至于直方图均衡化过的图像灰度级能跨越更大的范围。

在本节前面讨论了具有覆盖整个灰度范围的灰度级值的许多优点。除了产生有这样趋势的灰度级,刚才推导的方法还具有另一个优点,即它完全"自动化"。也就是说,已知一幅图像,直方图均衡化处理仅仅涉及执行式(3.3.8),该式基于能直接从已知图像提取的信息,不需要更多的参数说明。我们也注意到计算简单性是应用这一技术的要求。

从 s 回到 r 的反变换形式表示为:

$$r_k = T^{-1}(s_k) \quad k = 0,1,2,\cdots,L-1 \tag{3.3.9}$$

只要灰度级 $r_k(k=0,1,2,\cdots,L-1)$ 均出现于输入图像,就可看出(见习题3.9)式(3.3.9)的反变换满足此节前面给出的条件(a)和(b)。尽管反变换不用于直方图均衡化,但它在下节研究的直方图匹配方案中起核心作用。下节中还要详细讨论如何运用直方图处理技术。

例 3.3 直方图均衡化

图 3.17(a)显示了图 3.15 的四幅图像,图 3.17(b)显示了对这些图像中的每一个执行直方图均衡化后的结果。前三种结果(从上到下)显示了重要改进效果。如预期的那样,直方图均衡化在第四幅图像中没有在视觉上产生重要不同,这是因为这幅图像的直方图已扩展了全部灰度级范围。用于产生图 3.17(b)的变换函数被示于图 3.18。这些函数由原始图像[见图 3.15(b)]经式(3.3.8)产生。注意,变换(4)有一基本线性形状,再次表明第四幅输入图像灰度级几乎为均匀分布。如刚才所说的,我们希望直方图均衡化在这种情况下对图像表现没有影响。

均衡化后的图像直方图示于图 3.17(c)。有趣的是,当这些直方图不同时,直方图均衡化后的图像本身在视觉上很相似。但这并非不希望的,因为左侧的一列图像之间的差异仅仅是一种简单对比,而不是内容上的。也就是说,因为图像有相同的内容,从直方图均衡化导致的对比增强足以使得结果图像中,在视觉上难以区分灰度级的差别。在左侧一列的图像中给出了有意义的差别对比,这个例子说明了直方图均衡化作为适应性强的增强工具的强大作用。

3.3.2 直方图匹配(规定化)

如前所述,直方图均衡化能自动地确定变换函数,该函数寻求产生有均匀直方图的输出图像。当需要自动增强时,这是一个好方法,因为由这种技术得到的结果可预知,并且这种方法操作简单。我们在本节中还指出,有一些应用用均匀直方图的基本增强并不是最好的方法。尤其是,有时可以指定希望处理的图像所具有的直方图形状。这种用于产生处理后有特殊直方图的图像的方法,叫做直方图匹配或直方图规定化处理。

方法的推导

让我们回到连续灰度级 r 和 z(看做连续随机变量),令 $P_r(r)$ 和 $P_z(z)$ 为它们对应的连续概率密度函数。在这里,r 和 z 分别代表输入和输出(已处理)图像的灰度级。从输入图像估计 $P_r(r)$,而 $P_z(z)$ 为希望输出图像具有的规定概率密度函数。

令 s 为一随机变量,且有:

$$s = T(r) = \int_0^r p_r(w)\mathrm{d}w \tag{3.3.10}$$

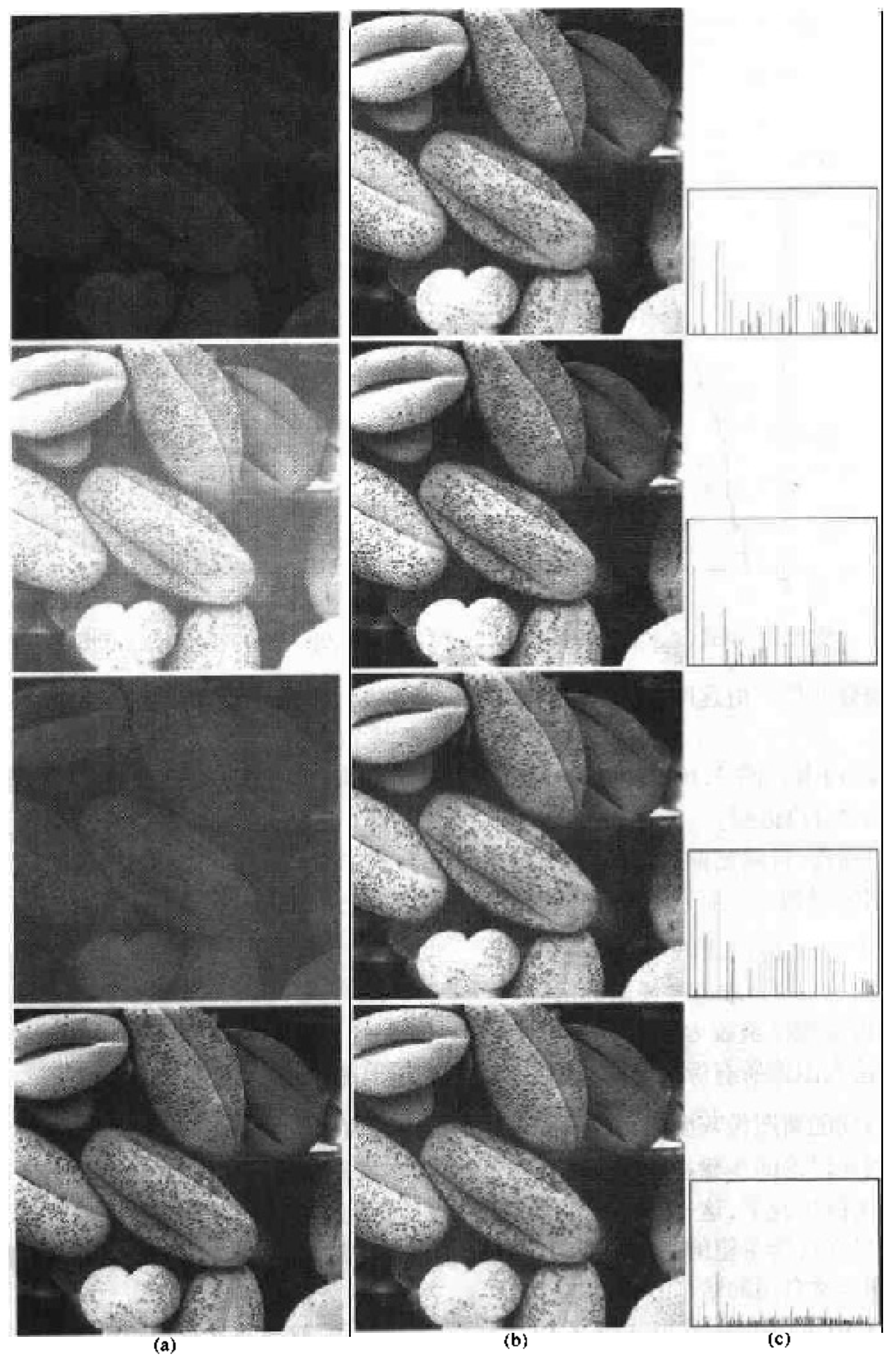


图 3.17　(a)图 3.15 的图像,(b)直方图均衡化的结果,(c)相应的直方图

其中 w 为积分变量。我们发现这个表达式为式(3.3.4)直方图均衡化的连续形式。然后假设定义随机变量 z,且有:

$$G(z) = \int_0^z p_z(t)\mathrm{d}t = s \tag{3.3.11}$$

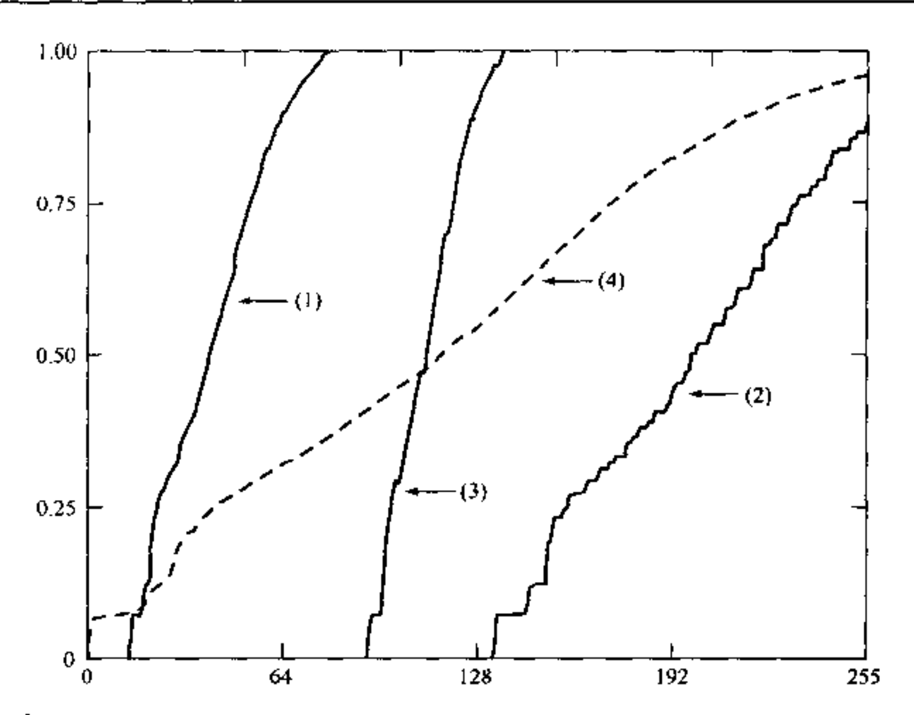


图 3.18 用式(3.3.8),从图 3.17(a)中的图像直方图得到的变换函数(1)到(4)

其中 t 为积分变量。由这两个等式可得到 $G(z)=T(r)$,因此,z 必须满足条件:

$$z = G^{-1}(s) = G^{-1}[T(r)] \tag{3.3.12}$$

变换函数 $T(r)$由式(3.3.10)得到,$p_r(r)$由输入图像估值。类似地,变换函数 $G(z)$因 $p_z(z)$已知而由式(3.3.11)得到。

设 G^{-1}存在,且满足前一节所述条件(a)和(b),式(3.3.10)到式(3.3.12)说明用下列步骤可由输入图像得到一个有规定概率密度函数的图像:

(1)由式(3.3.10)求得变换函数 $T(r)$。

(2)由式(3.3.11)求得变换函数 $G(z)$。

(3)求得反变换函数 G^{-1}。

(4)对输入图像所有像素应用式(3.3.12)得到输出图像。

以上步骤得到的新图像灰度级具有事先规定的概率密度函数 $p_z(z)$。

尽管刚才讨论的步骤在理论上可直接进行,但在实践中得到 $T(r)$和 G^{-1}却不太可能。幸运的是,在离散情况下,这一问题在相当大的程度上被简化了。付出的代价与直方图均衡化是相同的,这里仅仅所希望的直方图近似是可以获得的。然而,尽管如此,即使用很粗糙的近似也可以得到非常有用的结果。

式(3.3.10)的离散公式由式(3.3.8)给出,为便于学习,这里再次列出:

$$\begin{aligned} s_k = T(r_k) &= \sum_{j=0}^{k} p_r(r_j) \\ &= \sum_{j=0}^{k} \frac{n_j}{n} \quad k = 0,1,2,\cdots,L-1 \end{aligned} \tag{3.3.13}$$

其中 n 为图像中像素的数量和,n_j 为灰度级为 r_j 的像素数量,L 为离散灰度级的数量。类似地,式(3.3.11)的离散表达式由给定的直方图 $p_z(z_i)$($i=0,1,2,\cdots,L-1$)得到,且有形式:

$$v_k = G(z_k) = \sum_{i=0}^{k} p_z(z_i) = s_k \quad k = 0,1,2,\cdots,L-1 \tag{3.3.14}$$

如连续情况，我们寻找满足等式的 z 值。为后续讨论清楚起见，变量 v_k 附在这里。最后，式(3.3.12)的离散形式由下式给出：

$$z_k = G^{-1}[T(r_k)] \quad k = 0,1,2,\cdots,L-1 \tag{3.3.15}$$

或由式(3.3.13)得到：

$$z_k = G^{-1}(s_k) \quad k = 0,1,2,\cdots,L-1 \tag{3.3.16}$$

式(3.3.13)到式(3.3.16)是数字图像直方图匹配的基本公式。式(3.3.13)是基于原始图像直方图，从原始图像灰度级到对应灰度级 s_k 的映射，该原始图像的直方图从图像的像素计算得到。式(3.3.14)从给定的直方图 $p_z(z)$ 计算变换函数 G。最后，式(3.3.15)或等价的式(3.3.16)给出了此直方图的图像所希望的灰度级(的近似)。前两式因其量均已知而很容易实现。式(3.3.16)可直接实现，但需要附加说明。

实现

我们由以下说明开始：(1)灰度级 $\{r_j\}$，$\{s_j\}$ 和 $\{z_j\}$ $(j = 0,1,2,\cdots,L-1)$ 的每一个集合为 $L\times 1$维的一维阵列；(2)所有从 r 到 s 和从 s 到 z 的映射都是给定像素值和这些阵列间的简单查表过程；(3)这些阵列的每个元素(如 s_k)包括两条重要信息，即下标 k 表示元素在阵列中的位置，s 表示该位置的值；(4)只关心整数像素值。例如，在一个 8 比特图像中，$L = 256$，且阵列中每个元素所涉及的范围为 0 到 255 的整数。这就是说，现在处理的灰度级在区间$[0, L-1]$，而不是前面用于简化直方图处理技术的归一化区间[0,1]。

为了知道直方图匹配实际上如何进行，考虑图 3.19(a)并暂时忽略此图和图 3.19(c)的联系。图 3.19(a)显示了一假设的由给定图像得到的离散变换函数 $s = T(r)$。图像中第一个灰度级 r_1 映射到 s_1，第二个灰度级 r_2 映射到 s_2，第 k 个灰度级 r_k 映射到 s_k，等等(这里最重要的一点是这些值之间的对应顺序)。阵列中的每一个值 s_j 用式(3.3.13)预先进行计算，因此，映射处理仅仅使用像素的实际值作为阵列中的索引来决定相应的 s 值。这个过程非常简单，因为我们处理的是整数。例如，对于 8 比特像素其值为 127 的映射 s，将在阵列$\{s_j\}$中可能的 256 个位置中的第 128 个位置找到(记住，我们是从 0 开始的)。如果我们现在停止，并为输入图像的每一个像素值用上述方法进行映射，根据式(3.3.8)，将输出直方图均衡化了的图像。

为了执行直方图匹配，必须再进行一步操作。图 3.19(b)是一个由给定直方图 $P_z(z)$用式(3.3.14)得到的一个假设的变换函数 G。对每一个 z_q，这个变换函数得到一个对应的值 v_q。这个映射由图 3.19(b)的箭头符示出。相反，若已知任意 v_q 值，可以由 G^{-1}得到对应的值 z_q。根据这幅图，我们可以倒转箭头符方向而由 v_q 映射到z_q。不管怎样，由式(3.3.14)的定义知道由对应的下标有 $v = s$，因此，可以确切地利用这种处理找到任意对应于 s_k 的值 z_k，z_k 是从等式 $s_k = T(r_k)$计算得来的。这一思想示于图 3.19(c)。

因为的确不知道 z 的值(找到这些值恰好是直方图匹配的目标)，所以必须借用某种迭代方案由 s 找到z。事实上，由于处理的是整数使得这一过程变得尤为简便。从根本上说，因为 $v_k = s_k$，由式(3.3.14)可知，所找的 z 值必须满足等式 $G(z_k) = s_k$，或者$G(z_k) - s_k = 0$。

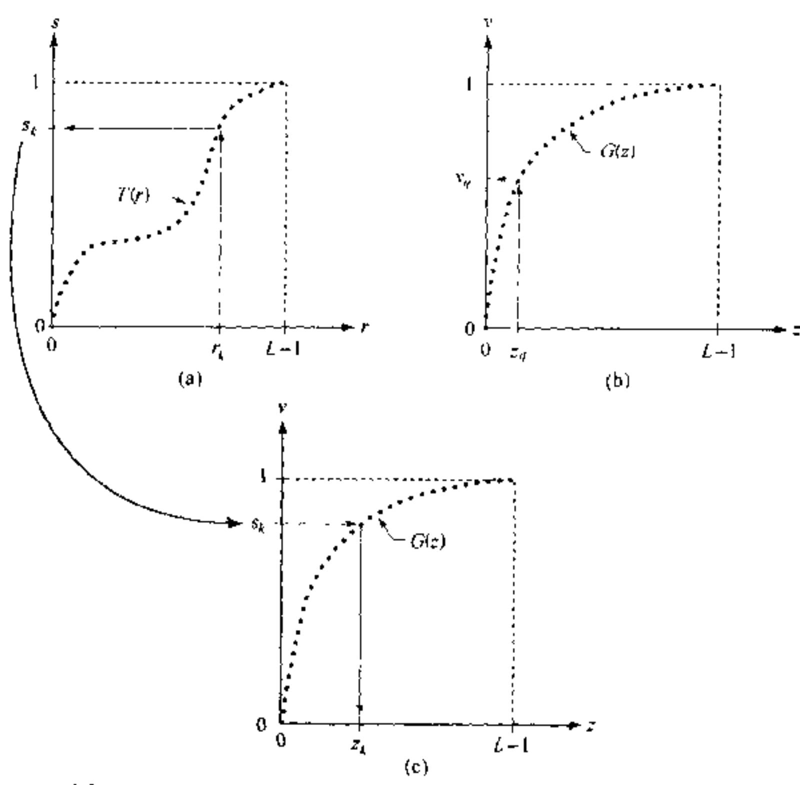


图 3.19 (a)经 $T(r)$从 r_k 到s_k 映射的图解,(b)经 $G(z)$从 z_q 映射到其相应的值v_q,(c)从s_k到其相应的值 z_k的逆映射

因此,要找到 s_k 所对应的值z_k,只需在 z 值上迭代,以满足等式(此时,$k=0,1,2,\cdots,L-1$)。式(3.3.16)也一样,除了不需要求 G 的反函数,这是因为我们正对 z 迭代。因为处理的是整数,所能得到的满足等式$(G(z_k)-s_k)=0$最接近的整数,对每一个 k 值使$z_k=\hat{z}$,其中 $\hat{z}$ 是区间$[0,L-1]$中最小的整数,即:

$$(G(\hat{z})-s_k)\geqslant 0 \quad k=0,1,2,\cdots,L-1 \tag{3.3.17}$$

给定一个值 s_k,关于图 3.19(c)概念上应从 $\hat{z}=0$ 开始,且以整数步长逐渐增加至满足式(3.3.17),在该点上,我们令 $z_k=\hat{z}$。重复这一过程,每一个 k 值都得到所需要的从 s 到 z 的映射,这就实现了式(3.3.16)。事实上,我们不必每次都从 $\hat{z}=0$ 开始,因为 s_k 值已知是单调递增的。因此,对于 $k=k+1$,可以从 $\hat{z}=z_k$ 开始,并以整数递增。

对刚才实现的直方图匹配的操作步骤可总结如下:

1. 求出已知图像的直方图。
2. 利用式(3.3.13)对每一灰度级 r_k 预计算映射灰度级 s_k。
3. 利用式(3.3.14)从给定的 $P_z(z)$得到变换函数 G。
4. 利用式(3.3.17)定义的迭代方案对每一个 s_k 值预计算值。
5. 对于原始图像的每个像素,若像素值为 r_k,将该值映射到其对应的灰度级 s_k;然后映射灰度级 s_k 到最终灰度级 z_k。

利用步骤(2)和步骤(4)的预计算值进行映射。注意步骤(5)对被处理图像的每个像素进行了两次映射。第一次映射仅仅是直方图均衡化。如果经直方图均衡化了的图像不是所要求的,

显然，为了节省一次中间步骤，将两次变换结合为一次是有利的。

最后，注意到，甚至在离散情况下，都需要考虑 G^{-1} 要满足前一节中的条件(a)和(b)。不难看出(习题3.9)，保证 G^{-1} 为单值且单调的惟一方法是要求 G 严格单调(也就是一直增加)，这仅仅意味着规定化直方图 $P_z(z_i)$ 在式(3.3.14)中没有零值。

例3.4 直方图均衡化和直方图匹配间的比较

图3.20(a)显示了火星的卫星(Phobos)图像，该图像由NASA的表面探测器拍摄。图3.20(b)显示了图3.20(a)的直方图。图像主要是大片的暗区域，产生了像素集中于灰度级暗端的直方图特点。猛一看，会觉得直方图均衡化是图像增强的一个好办法，使暗区域的细节更清楚。而下面的讨论将证明事实并非如此。

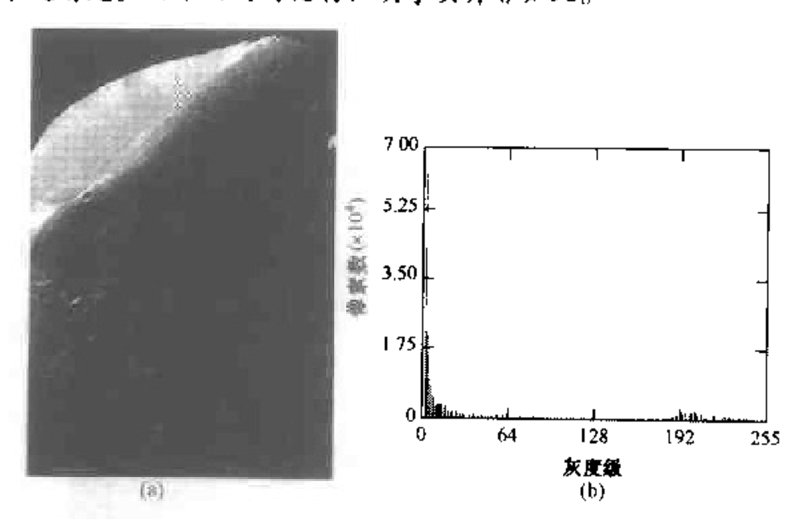


图3.20 (a)由NASA表面探测器拍摄的火星卫星图像，(b)直方图(原图像由NASA提供)

图3.21(a)显示了由图3.20(b)所示的直方图得到的直方图均衡化变换[式(3.3.8)或式(3.3.13)]。这个变换函数最相关的特性是快速地从灰度0升到接近190。这是因为输入直方图中有大量的像素灰度很接近0。当这种变换应用于输入图像的灰度以获得直方图的均衡效果时，净效果是把非常窄的暗像素区间映射到输出图像灰度级的高端。因为输入图像的大量像素在此区间有精确的灰度值，我们希望图像能够有一个亮的、看起来"冲淡"了的外观。如图3.21(b)所示，情况就是如此，这个图像的直方图示于图3.21(c)，注意，所有灰度级基本上都偏向了灰度范围的上半部。

由于图3.21(a)中变换函数的问题是由原图中大量灰度接近于0的像素引起的，故以一种合理的方法修改图像的直方图，以便使其不再有这一属性。图3.22(a)显示了一个人为指定的函数，它保持了原有直方图的形状，但在图像灰度暗区的灰度值却有一个平滑的变化。对这个函数取样为256个等间隔的离散值，以产生所需的规定化直方图。使用式(3.3.14)从这一直方图得到的变换函数 $G(z)$，在图3.22(b)中标注为变换(1)。类似地，由式(3.3.16)导出的反变换函数在图3.22(b)中标注为变换(2)[这个函数是通过讨论过的关于式(3.3.17)的迭代技术得到的]。图3.22(c)中增强的图像是对图3.21(b)中直方图均衡化后的像素用变换(2)得到的。比较这两幅图像可以看出，在用直方图均衡化得到的结果的基础上，对直方图规定化处理的改进是很有意义的。注意到，原始直方图适度

的改变对在图像增强中得到有效的改进是很有必要的。图 3.22(c)的直方图示于图 3.22(d)。这个直方图最容易辨别的特征是低端如何根据需要移到较亮的右端。

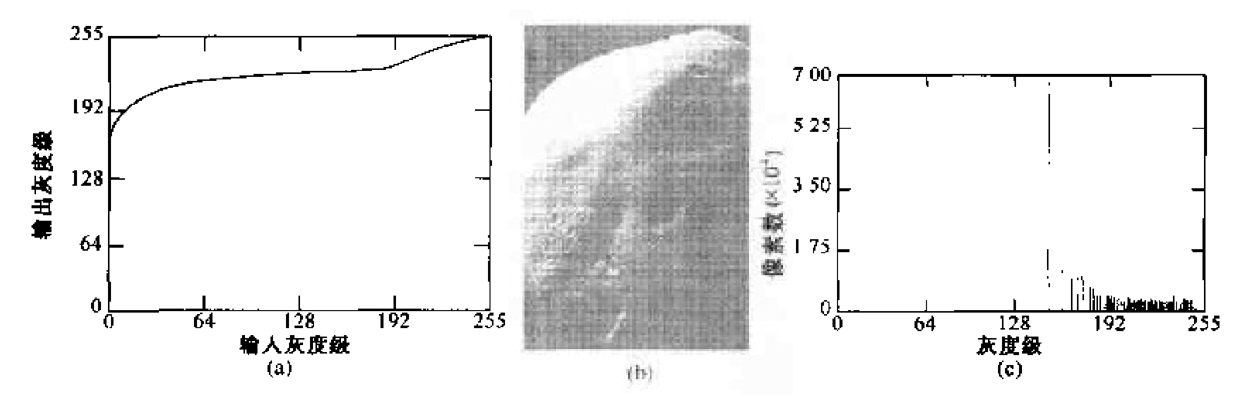


图 3.21 (a)直方图均衡化变换函数,(b)直方图均衡过的图像(注意,类似"冲洗"过的外观),(c)(b)的直方图

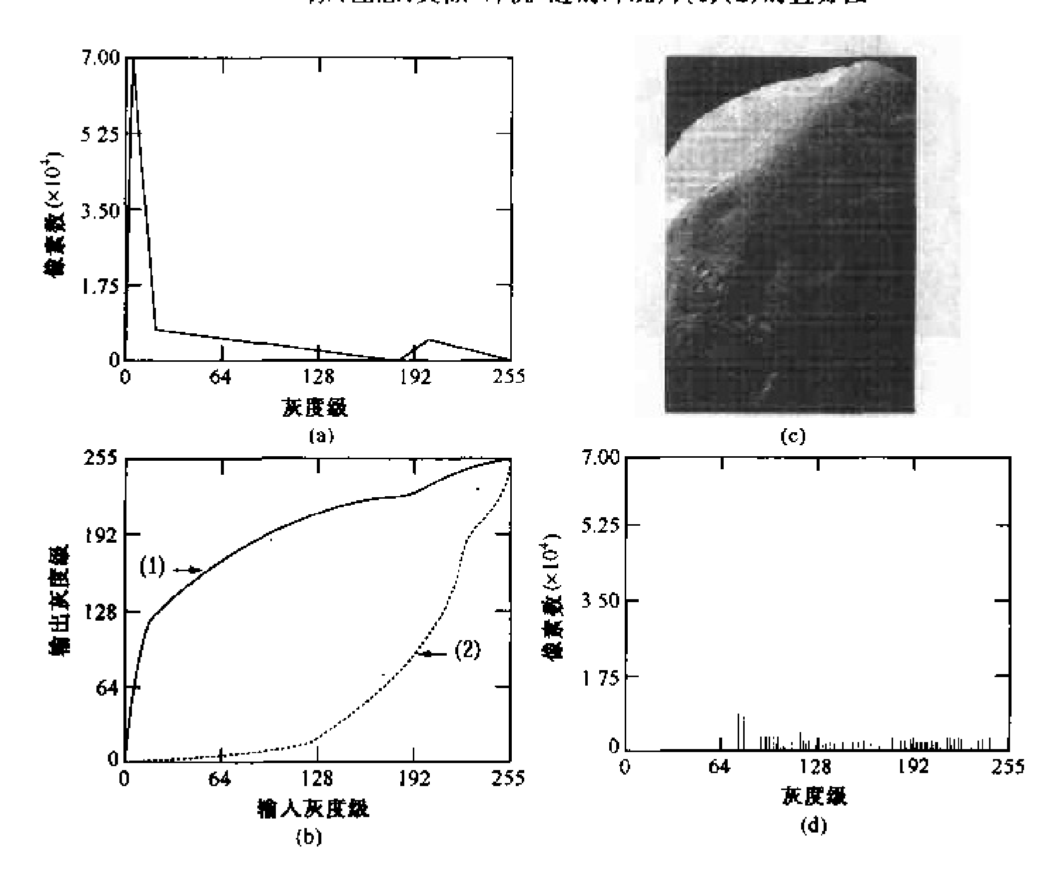


图 3.22 (a)规定的直方图,(b)曲线(1)是用(a)中直方图从式(3.3.14)得到的,曲线(2)用式(3.3.17)迭代得到,(c)用曲线(2)映射增强的图像,(d)(c)的直方图

虽然到现在为止对直方图均衡可能已很清楚了,但我们在结束这一部分之前要强调直方图规定化在大多数时候都是试凑过程。有时有可能会遇到这样的情况,即定义"平均"直方图的外观,并把它用做规定化的直方图。在这种情况下,直方图规定化变成了直接处理。不管怎样,总的来说,并没有规定化直方图的规定,对于任何一个给定的增强任务都必须借助于实际分析。

3.3.3 局部增强

前两节讨论的直方图处理方法是全局性的,在某种意义上,像素是被基于整幅图像灰度满意度的变换函数所修改的。这个全局方法适用整个图像的增强,但有时对图像小区域细节的局部增强也仍然是适用的。在这些区域中像素数在全局变换的计算中可能被忽略,因为它们没有必要确保局部增强。解决的方法就是在图像中每一个像素的邻域中,根据灰度级分布(或其他特性)设计变换函数。虽然基于邻域的处理方法是 3.5 节的主题,但在这里讨论局部直方图处理是为了内容的清楚和完整性。这样读者在后续讨论中不会遇到什么困难。

以前描述的直方图处理技术很容易适应局部增强,该过程定义一个方形或矩形的邻域并把该区域的中心从一像素移至另一像素。在每一个位置的邻域中该点的直方图都要被计算,并且得到的不是直方图均衡化就是规定化变换函数。这个函数最终被用来映射邻域中心像素的灰度。相邻区域的中心然后被移至相邻像素位置并重复这个处理过程。当对某区域进行逐像素转移时,由于只有邻域中新的一行或一列改变,所以可以在每一步移动中,以新数据更新前一个位置获得的直方图(习题 3.11)。这种方法相比邻域每移动一个像素就对基于所有像素的直方图进行计算,有明显的优点。有时使用非重叠区域是减少计算量的另一种方法。但是这种方法通常会出现不希望的棋盘效果。

例 3.5　局部直方图增强

图 3.23(a)显示了一幅被轻度模糊以减少噪声的图像(参看 3.6.1 节关于模糊的内容)。图 3.23(b)显示了全局直方图均衡的结果。当这种技术用于平滑噪声区域时,情况经常如此,如图 3.23(b)所示,对比度稍微提高,噪声就会明显地增强。可以说,这种方法没有带来新的结构性细节。然而,局部直方图均衡使用了一个 7×7 的邻域以显示在大的暗方形中的小方形。大小方形在灰度上很接近,但小方形的尺寸太小,故对全局直方图均衡化的影响不大。也应注意到图 3.23(c)中有细的噪声纹理,这是在小邻域使用局部处理的结果。

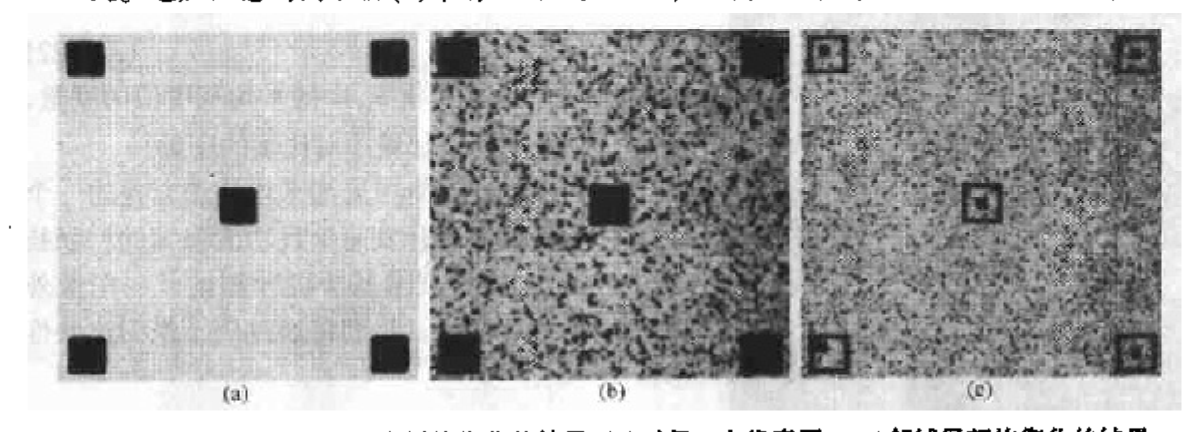


图 3.23　(a)原像,(b)全局均衡化的结果,(c)对每一个像素用 7×7 邻域局部均衡化的结果

3.3.4 在图像增强中使用直方图统计学

尽管可以直接使用直方图对图像进行增强,但是我们也可以使用直接从直方图获得的统计参数。令 r 表示在区间 $[0, L-1]$ 上代表离散灰度的离散随机变量,并且令 $p(r_i)$ 代表对应于 r 的第 i 个值的归一化直方图分量。正如在本节前面指出的,可以把 $p(r_i)$ 看做灰度级 r_i 出

现的概率估计值。r 的第 n 个矩的平均值定义如下：

$$\mu_n(r) = \sum_{i=0}^{L-1}(r_i - m)^n p(r_i) \tag{3.3.18}$$

此处 m 是 r 的平均值(其灰度级均值)：

$$m = \sum_{i=0}^{L-1} r_i p(r_i) \tag{3.3.19}$$

由式(3.3.18)和式(3.3.19)导出 $\mu_0 = 1, \mu_1 = 0$，其二阶矩的表达式如下：

$$\mu_2(r) = \sum_{i=0}^{L-1}(r_i - m)^2 p(r_i) \tag{3.3.20}$$

我们认为这个式子是 r 的方差，通常以 $\sigma^2(r)$ 表示。标准差被简单定义为方差的平方根。在第 11 章将讨论有关图像描述的矩。然而，就增强来说，最关心的是一幅图像灰度的平均值和方差(或标准差)(它是平均对比度的量度)。

对图像增强，我们考虑平均值和方差的两种用途。全局平均值和方差是对整幅图像进行度量，并是对整幅图像强度和对比度的初步粗调整。这两种方法更强大的应用是在局部增强中，这里局部平均值和方差被用做实施改变的基础，而这种改变依靠图像中对每个像素预先定义的区域的图像特性。

令 (x, y) 为某一图像中像素的坐标，令 S_{xy} 表示一确定大小的邻域(子图像)，其中心在 (x, y)。根据式(3.3.19)，在 S_{xy} 中像素的平均值 $m_{S_{xy}}$ 能以下式计算：

$$m_{S_{xy}} = \sum_{(s,t)\in S_{xy}} r_{s,t} p(r_{s,t}) \tag{3.3.21}$$

此处 $r_{s,t}$ 是在邻域中坐标 (s, t) 处的灰度，且 $p(r_{s,t})$ 是与灰度值对应的邻域归一化直方图分量。类似地，在式(3.3.20)中，区域 S_{xy} 中像素的灰度级方差由下式给出：

$$\sigma^2_{S_{xy}} = \sum_{(s,t)\in S_{xy}} [r_{s,t} - m_{S_{xy}}]^2 p(r_{s,t}) \tag{3.3.22}$$

局部平均值是对邻域 S_{xy} 中的平均灰度值的度量，方差(或标准差)是邻域中对比度的度量。

在图像处理中使用局部平均值和方差的一个重要特点是提供了开发简单且功能强大的增强技术的灵活性，这种技术基于可预测的且与图像外观相近的统计度量。我们将通过例子说明这些性质。

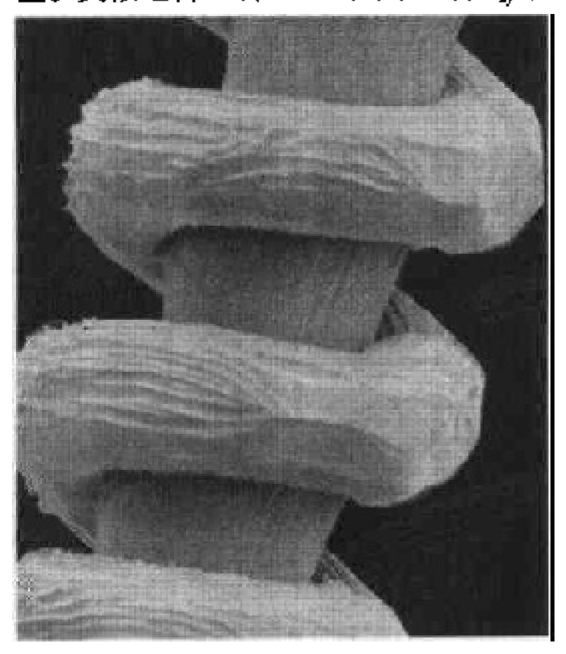


图 3.24　放大约 130 倍的钨丝和支架图像(原图像由俄勒冈大学地质科学系的Michael Shaffer先生提供)

例 3.6　基于局部统计的增强

图 3.24 显示了一根绕在支架上的钨丝 SEM(扫描电子显微镜)图像。在图像中央的钨丝和其支架很清楚并很容易分析。在图像的右边有另一根钨丝的结构，但是它更暗，并且大小和特征并不清楚。通过对比操作进行的局部增强在这里是很理

想的方法,这时图像的一部分是可接受的,其他部分则可能包含我们所关心的隐藏特征。在这个特殊例子中,问题是增强暗的部分,而对亮的部分尽量不动,因为它并不需要增强。可以使用本节所提供的概念阐明一种增强方法,而这种方法能分辨暗与亮的不同,同时只增强暗的部分。判断一个点(x,y)是暗还是亮的方法是把局部平均灰度和图像的平均灰度进行对比,后者称之为全局灰度并用M_G表示。后者的数量是通过令S包含整幅图像而得到的。因此,增强方案的首要因素是:如果$m_{s_{xy}} \leqslant k_0 M_G$,则认为像素在点$(x,y)$可作为候选处理对象,此处$k_0$是一个小于1.0的正常数。由于我们感兴趣的增强区域有较低的对比度,还需要一个测量方法决定一个区域的对比度是否是要进行增强的候选点。因此,如果$\sigma_{s_{xy}} \leqslant k_2 D_G$,我们认为在点$(x,y)$处的像素是增强候选点,此处$D_G$是全局标准差,且$k_2$为正常数。如果我们增强亮区,这个常数将大于1.0,对于暗区则小于1.0。最后,我们需要限制能够接受的对比度最低值,否则该过程甚至可能增强标准差为0的恒定区域。因此,我们通过要求$k_1 D_G \leqslant \sigma_{s_{xy}}$,$k_1 < k_2$,对局部标准差设置一个较低的限制值。满足局部增强所有条件的一个(x,y)处的像素简单地通过乘以一个固定常数E来处理,以增加(或减少)与图像其余部分相关的灰度级。不满足增强条件的像素值不变。

增强的方法可总结如下:令$f(x,y)$表示在图像任何坐标(x,y)处的像素值,且令$g(x,y)$代表这些坐标上相应增强的像素。

$$g(x,y)=\begin{cases} E\cdot f(x,y) & \text{如果 } m_{s_{xy}} \leqslant k_0 M_G \text{ 且 } k_1 D_G \leqslant \sigma_{s_{xy}} \leqslant k_2 D_G \\ f(x,y) & \text{其他} \end{cases}$$

此处,如前所述的,E,k_0,k_1和k_2是特定的参数;M_G是输入图像的全局平均值;D_G是全局标准差。

通常,成功选择一个参数需要做一些试验来熟悉一种给定的图像或一类图像。在这种情况下,常常选择下列数值:$E=4.0$,$k_0=0.4$,$k_1=0.02$和$k_2=0.4$。对E选择相对较低的值0.4,在一个欲增强的区域(该区域较暗),用该区域的灰度乘以它时,结果将始终趋于灰度级的暗端,并保留了图像的视觉平衡。很显然,观察一幅图像时由于需要增强的区域确实比全局平均值的一半还要暗,故选k_0比全局平均值的一半略小。类似的分析导出了k_1和k_2值的选择结果。总的说来,这些常数的选择并非难事,但这些选择必须根据手边的增强问题进行逻辑分析后才能得到。最后,对局部区域大小的选择因为要考虑到保留细节和减少计算量而尽量选择小的尺寸。因此,我们选择3×3的小区域。

图3.25对所有的(x,y)值显示了$m_{S_{xy}}$的值。由于对每一个(x,y)的值,$m_{S_{xy}}$是中心在(x,y)的3×3领域内像素的平均值,我们预期结果会与原图像相似,但是,稍微有点模糊。这确实是如图3.25(a)所示的情况。图3.25(b)显示了用所有的$\sigma_{s_{xy}}$值形成的图像。类似地,我们可以在每一个坐标对(x,y)处,放大$f(x,y)$的值至范围之外形成$g(x,y)$而构造一幅图像。因为该值不是1就是E,所以图像如图3.25(c)所示是二值的。相对于暗区的是1,相时于亮区的是E。这样,在图3.25(c)中任何亮点意味着,在坐标对(x,y)处,增强过程用E乘以$f(x,y)$产生了一个增强的像素。暗点表示在该坐标处增强过程没有放大像素值。

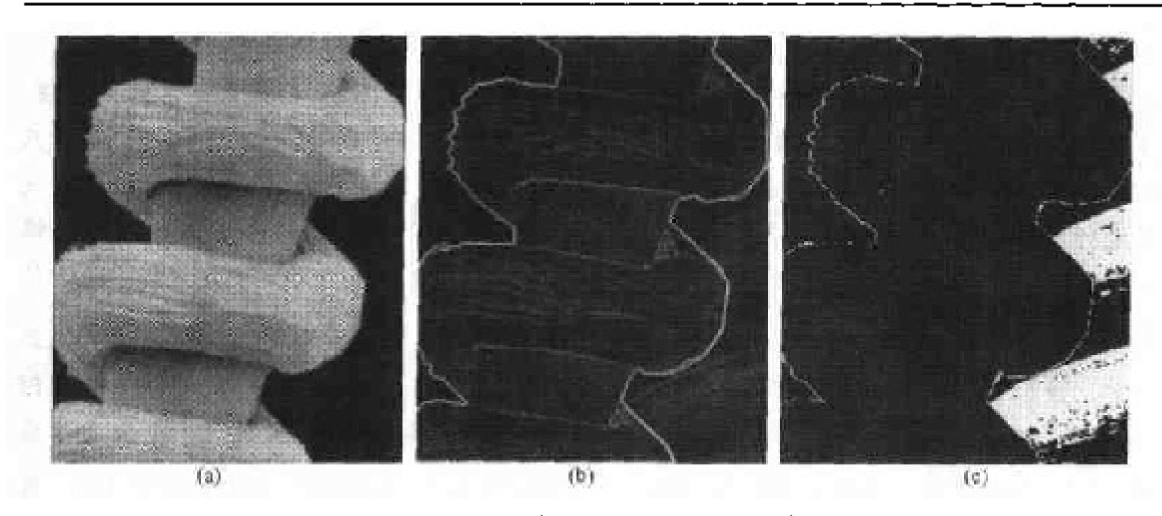


图 3.25 (a)从图 3.24 用式(3.3.21)得到的局部平均形成的图像;(b)从图 3.24 用式(3.3.22)通过局部标准差形成的图像;(c)用于产生图 3.26 所示增强图像的放大常数形成的图像

用刚刚讨论的方法增强后的图像示于图 3.26。用该图像与图 3.24 中的原图像相比,可以明显地看到增强后图像的右侧已经显示出的细节。它指出图像没有增强的部分(亮区)基本上被原封不动地保留了下来。我们还注意到在线圈与支架交接处的阴影区和灯丝与背景之间出现了小亮点。这些都是由增强引起的不希望的产物。换句话说,这些亮点满足了增强的准则并且其值被参数 E 放大了。噪声的引入是该方法固有的缺点,正如刚刚阐述的那样,这是非线性处理图像造成的。然而,在这里关键的一点是用大多数令人满意的方法增强了的图像确实突出了所希望的细节。



图 3.26 增强的 SEM 图像。与图 3.24 比较,特别注意在图像右侧增强的区域

不难想像刚才给出的例子中所用的多数方法对于局部增强的情况也适用。

3.4 用算术/逻辑操作增强

图像中的算术/逻辑操作主要以像素对像素为基础在两幅或多幅图像间进行(其中不包含逻辑"非"操作,它在单一影像中进行)。例如,两幅图像相减产生一幅新图像,这幅新图像在坐标(x,y)处的像素值与那两幅进行相减处理的图像中同一位置的像素值有所不同。通过使用硬件和软件,就可以实现对图像像素的算术/逻辑操作,这种操作可以一次处理一个点,也可以并行进行,即全部操作同时进行。

对图像的逻辑操作同样也是基于像素的。我们关心的只是"与"、"或"、"非"逻辑算子的实现,这三种逻辑算子完全是函数化的。换句话说,任何其他的逻辑算子都可以由这三个基本算子来实现。当我们对灰度级图像进行逻辑操作时,像素值作为一个二进制字符串来处理。例如,对一个 8 比特的黑色像素值(8 个比特都是 0 的串)进行"非"处理就会产生一个白色像素值(8 个比特都是 1 的串)。中间值也是用同样的方法处理得出的:将所有的 1 变为 0,反之亦然。另外,逻辑非算子执行与式(3.2.1)中的反比变换相同的功能。如图 3.27 所示的那样,"与"操作和"或"操作通常用做模板,即通过这些操作可以从一幅图像中提取子图像。在"与"和"或"图像模板中,亮的表示二进制码 1,黑的代表二进制码 0。模板处理有时可以作为一种感兴趣区(ROI)处理。就增强而言,模板主要用于分离要处理的区域,这时突出一个区域来区别图像的其他区域。此外,逻辑操作经常与第 9 章讨论的形态操作同时使用。

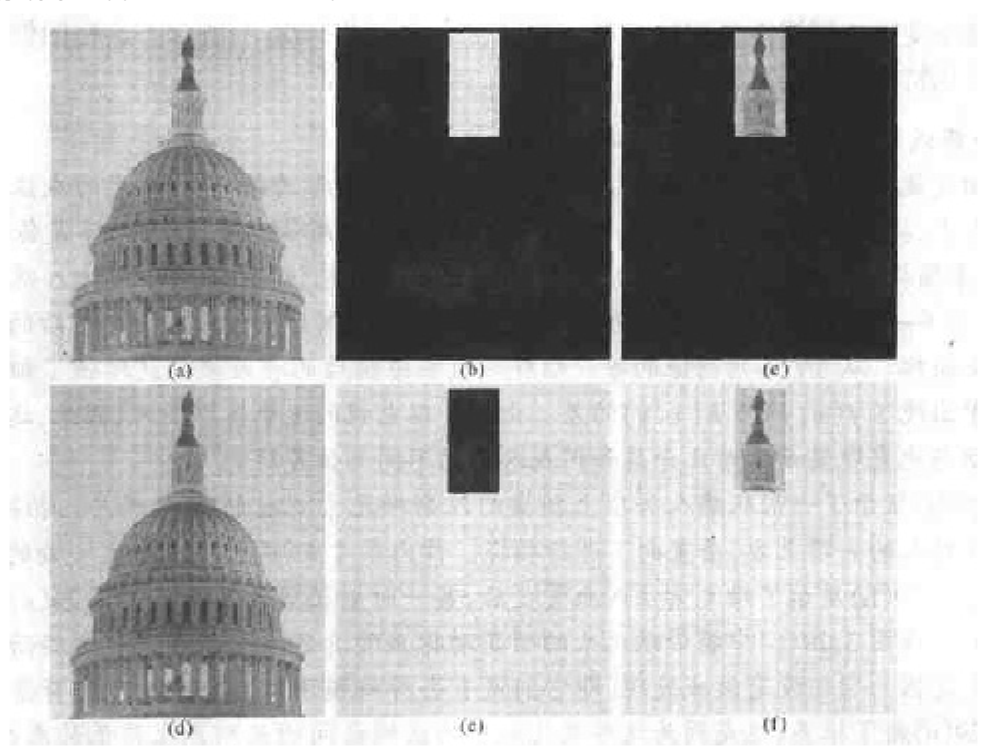


图 3.27　(a)原图像,(b)"与"图像模板,(c)对图像(a)和(b)进行"与"操作的结果,(d)原图像,(e)进行"或"图像模板,(f)对图像(d)和(e)进行"或"操作的结果

在四种算术操作中,减法与加法在图像增强处理中最为有用。我们简单地把两幅图像相除看做用一幅的取反图像与另一幅图像相乘。除了用一个常数与图像相乘以增加其平均灰度

的操作以外,图像乘法主要用于比前边讨论的逻辑模板处理更为广泛的模板操作增强处理。换句话说,用一幅图像去乘另一幅图像可直接用于灰度处理,而不仅仅是对二进码模板处理。我们在3.8节给出的例子中看到模板操作是一个强有力的工具。在本节的剩余部分中,将讨论基于减法和加法处理的图像增强技术,而图像乘法的其他应用技术将放在第5章中结合图像复原一起讨论。

3.4.1 图像减法处理

两幅图像 $f(x,y)$ 与 $h(x,y)$ 的差异表示为:

$$g(x,y) = f(x,y) - h(x,y) \tag{3.4.1}$$

图像的差异是通过计算这两幅图像所有对应像素点的差而得出的。减法处理最主要的作用就是增强两幅图像的差异。为了说明这个概念,我们简要地回顾3.2.4节中的讨论,其中说明了一幅图像的高阶比特面会携带大量的可见相关细节,低阶比特面则分布着一些细小的细节(通常是感觉不到的)。图3.28(a)显示了前面使用的为了说明比特面概念的分形图像。图3.28(b)显示了从原始图像中去除四个最后有效比特面(置0)的结果。这两幅图像在视觉上几乎完全一样,只是图3.28(b)中的灰度值存在极小的变化而使整个对比度稍微有所下降。这两幅图像对应像素间的差别示于图3.28(c)。可以看到,由于两幅图像之间的差异太小,以至于当用8比特显示时,差异图3.28(c)几乎是全黑的。为了能够解释得更清楚,我们可以对它进行如3.2节或3.3节讨论的对比度扩展变换处理。可以选择直方图均衡化的方法,但做一个近似幂率变换也同样能做到这一点。处理的结果如图3.28(d)所示。这幅图像很好地说明了将低阶比特面设置为0的效果。

例3.7 掩模式射线成像术中图像相减的应用

图像相减最成功和有益的商业应用之一是在医学领域称为掩模式X光的成像法。在这种技术中,掩模 $h(x,y)$ 是通过一台位于X光源对面的增强电视镜头(而不是传统的X光胶片)来捕捉病人的病变区域的X光图像。处理过程包括:将对比介质注入病人的血管之中,拍下一系列同一解剖区域的图像作为 $h(x,y)$,然后从注入对比介质后的相应图像中减去掩模。从TV序列图像的每一幅样片减去掩模后的净效果就是增强了细节的输出图像中出现了 $f(x,y)$ 与 $h(x,y)$ 的差。由于是以电视的速率抓取图片,因此,这一过程大体上动态地直观显示了对比介质如何在人体的不同部位蔓延。

图3.29(a)显示了一张从病人头顶上拍摄的注射碘元素之前的X光照片。拍摄时,摄像机放在病人的头顶上方,垂直向下进行拍摄。作为参考点,图像下三分之一处的亮点是脊柱。图3.29(b)显示了碘元素注入血管之后,某一时刻拍摄的图像与图3.29(a)的掩模间的差别。在图3.29(b)中携带碘元素的明亮动脉血管清楚地被增强了。这些动脉显得异常明亮是因为它们没有做减处理(即它们没有去掉掩模图像)。另外,所有背景区域要比图3.29(a)暗了很多,这是因为这些变化较小的区域之间的差别产生较低的灰度值,所以在差值图像中它呈现较暗的灰度。注意,在图3.29(a)中的脊柱较亮,而经过减处理后在图3.29(b)却变得很暗。

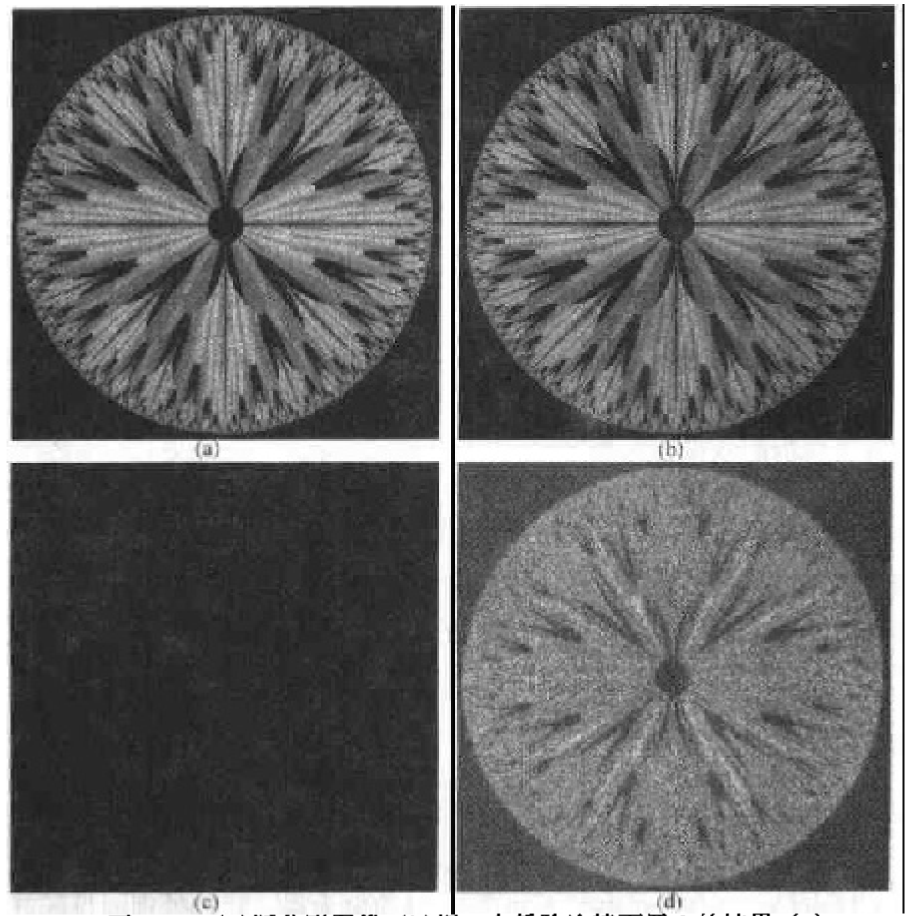


图 3.28　(a)原分形图像,(b)把 4 个低阶比特面置 0 的结果,(c)(a)和(b)间的差别,(d)直方图均衡后的差值图像(原图像由 Swarthmore 学院的 Melissa D. Binde 先生提供)

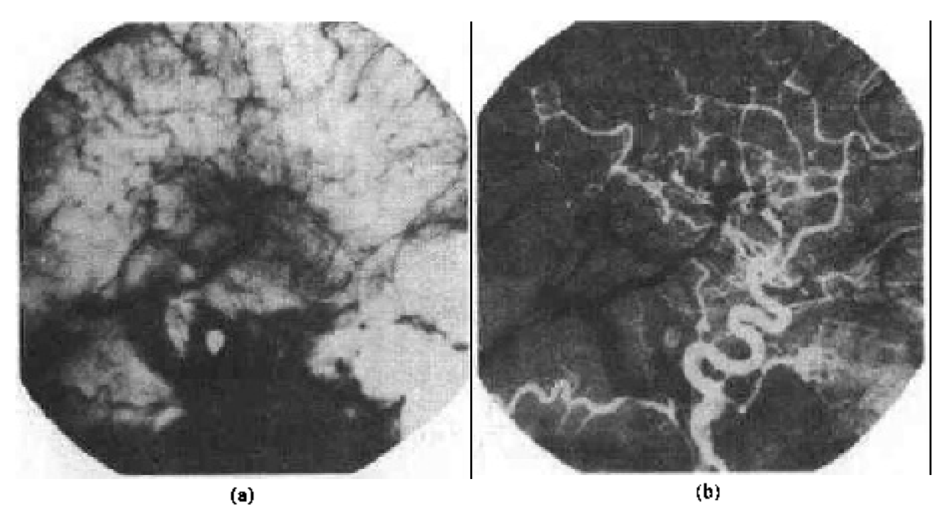


图 3.29　图像相减增强。(a)掩模图像,(b)减除掩模图像后的图像(把对比介质注入血管后拍摄的图像)

在结束这一节的讨论之前,做一点实践方面的评述。在实践中,大多数的图像由 8 位码显示(即使 24 比特的彩色图像也由 3 组 8 位码的通道组成),因此,像素值的大小不会超出 0 到

255的范围。在差值图像中,像素值的取值最小为 - 255,最大为255,因此,显示这一结果需要某种标度。有两种主要的方法标度差值图像:一种方法是对每个像素值再加255然后除以2。这种做法无法保证像素的取值可以覆盖0到255的全部8比特范围,但所有的像素值一定都在这一范围内。这种方法实现上快速而简单,但它也有一定的局限性,即整个显示范围没有得到充分利用,潜在的缺点是,在除2过程中固有的截尾误差通常将导致精确度的损失。

如果希望更高的精确度并使像素取值覆盖整个8比特的范围,可以采用另一种方法。首先,提取最小差值,并且把它的负值加到所有差值图像的像素中(这样就可以创作出一幅最小像素值为零的改进的差值图像)。然后,通过用255/Max值去乘每个像素(其中Max为改进的差值图像中最大像素取值)将图像中的所有像素标定到0到255的范围中。很明显,这种方法比起前一种来更为复杂而且难以实现。

在本节结束之前,我们还要提示一点,即通过图像减法处理的变化检测在图像分割领域也有另一方面的主要应用(图像分割是第10章的话题)。基本上,分割技术试图将一幅图像分成基于特定准则的区域。当准则"变化"时,对于图像分割就采用图像减处理。例如,在图像序列中跟踪(分割)行驶的车辆时,减法处理用来移去图像中那些静止的部分,剩余的只是图像中移动的元素及加性噪声。

3.4.2 图像平均处理

考虑一幅将噪声 $\eta(x,y)$ 加入到原始图像 $f(x,y)$ 形成的带有噪声的图像 $g(x,y)$,即:

$$g(x,y) = f(x,y) + \eta(x,y) \tag{3.4.2}$$

这里假设每个坐标点 (x,y) 上的噪声都不相关[①]且均值为零。我们处理的目标就是通过人为加入一系列噪声图像 $\{g_i(x,y)\}$ 来减少噪声。

如果噪声符合上述限制,会得到(问题3.15)如果对 K 幅不同的噪声图像取平均形成图像 $\bar{g}(x,y)$:

$$\bar{g}(x,y) = \frac{1}{K}\sum_{i=1}^{K} g_i(x,y) \tag{3.4.3}$$

则:

$$E\{\bar{g}(x,y)\} = f(x,y) \tag{3.4.4}$$

$$\sigma^2_{\bar{g}(x,y)} = \frac{1}{K}\sigma^2_{\eta(x,y)} \tag{3.4.5}$$

其中,在所有坐标点 (x,y) 上,$E\{\bar{g}(x,y)\}$ 是 $\bar{g}$ 的期望值,$\sigma^2_{\bar{g}(x,y)}$ 与 $\sigma^2_{\eta(x,y)}$ 分别是 $\bar{g}$ 与 η 的方差。在平均图像中任何一点的标准差为:

$$\sigma_{\bar{g}(x,y)} = \frac{1}{\sqrt{K}}\sigma_{\eta(x,y)} \tag{3.4.6}$$

当 K 增加时,式(3.4.5)和式(3.4.6)指出,在各个 (x,y) 位置上像素值的噪声变化率将减小。因为 $E\{\bar{g}(x,y)\} = f(x,y)$,这就意味着随着在图像均值处理中噪声图像使用量的增加,$\bar{g}(x,y)$ 越来越趋近于 $f(x,y)$。在实际应用中,为了防止在输出图像中引入模糊及其他人为

① 均值为 m 的随机变量 x 的方差定义为 $E[(x-m)^2]$,这里 $E\{\cdot\}$ 是宗量的期望值。两个随机变量 x_i 和 x_j 的协方差定义为 $E[(x_i-m_i)(x_j-m_j)]$。如果变量是不相关的,则它们的协方差为0。

影响，图像 $g_i(x,y)$ 必须被配准。

例3.8　用图像平均减少噪声

图像平均处理的一个重要应用是在天文学领域，在天文学中经常用极低亮度水平成像，这经常导致传感器噪声混入单幅图像，从而实际上使图像在分析时无法使用。图3.30(a)显示了由NASA的带有宽视野行星摄像机的Hubble太空望远镜拍摄到的一幅星系图，我们称之为NGC3314。NGC3314距地球140亿光年，其指向南半球Hydra星座方向。在星系中心附近形成环形的闪亮星体由星际气体与粉尘组成。图3.30(b)显示了相同的图像，只是受到了均值为0且标准差有64级灰度的非相关高斯噪声的干扰，因此，这幅图像对于特殊目的是不能用的。图3.30(c)到(f)分别是对8,16,64,128幅图像取平均后的结果。可以看到，取 $K=128$ 进行平均处理后的图像在视觉上已经与原始图像极为相似。

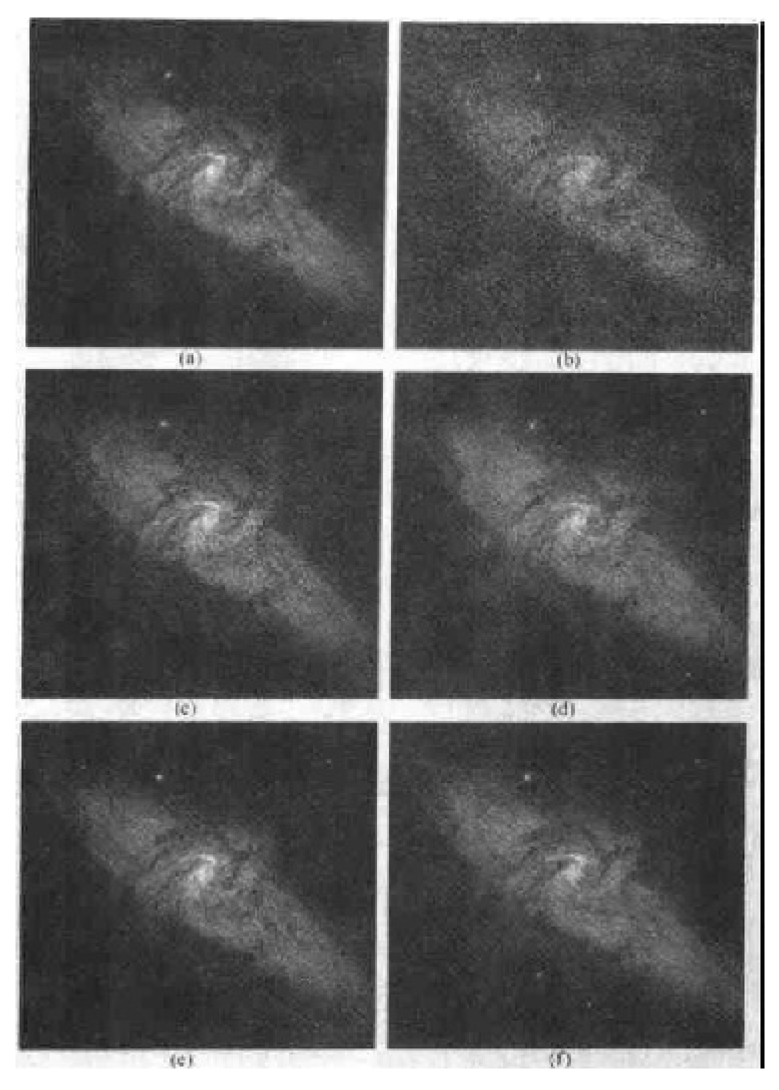
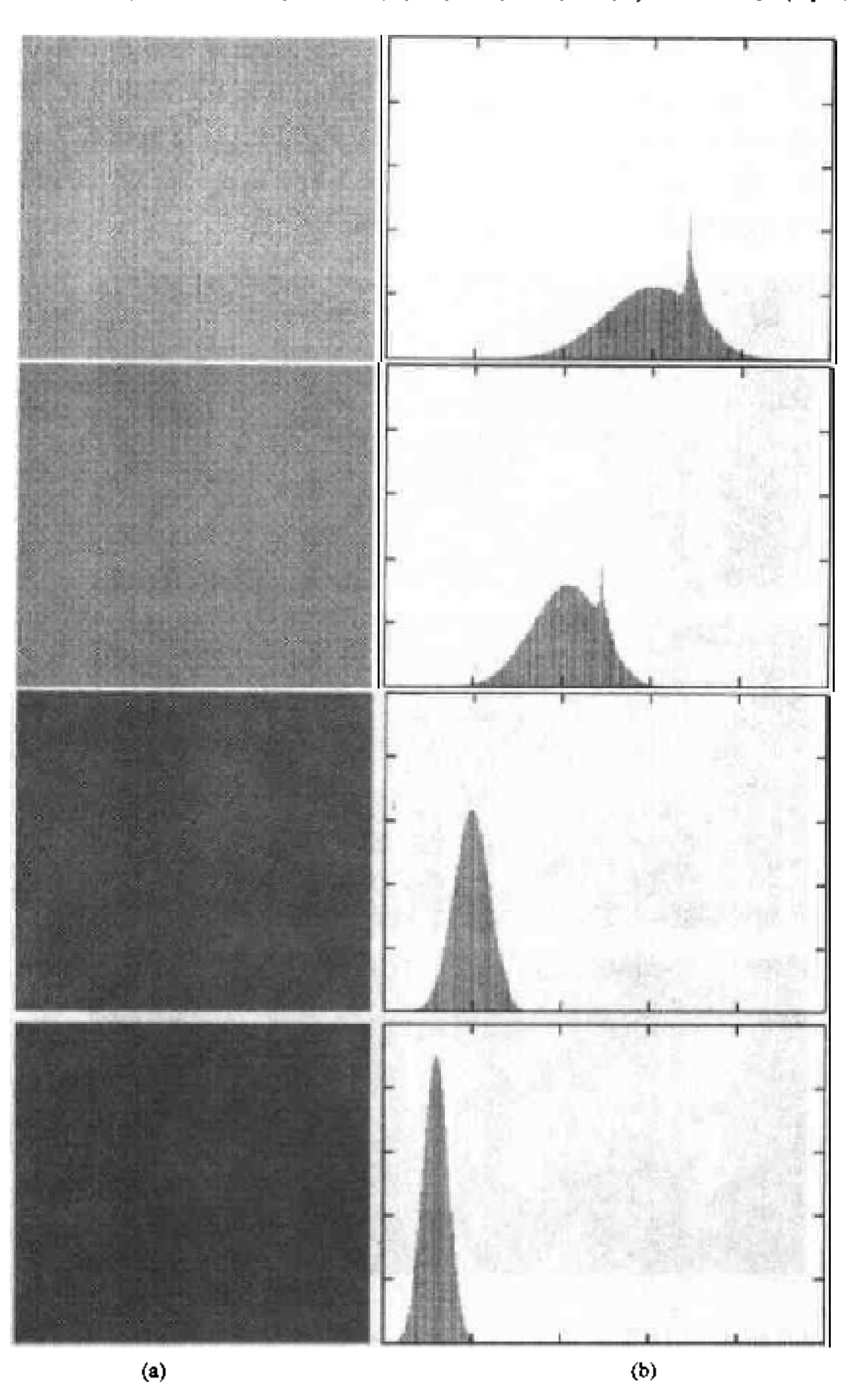


图3.30　(a) NGC 3314星团对图像，(b) 由0均值和64级灰度标准差的加性高斯噪声污染了的图像，(c)到(f)是 $K=8,16,64$ 和128的带噪声图像取平均的结果(原图像由NASA提供)

从图3.31我们可以更好地体会到随着 K 函数的增加，视觉噪声减少的现象是怎样产生的。该图像也显示了原始图像[图3.30(a)]与图3.30(c)到(f)中每一幅均值处理后图像

的差别。图中还显示了差值图像所对应的直方图。通常,直方图的垂直比例尺代表了像素的数目,其范围在[0,2.6×10^4],而水平比例尺代表了灰度级,其范围在[0,255]。我们注意到在直方图中,随着 K 的增加差值图像的均值和标准差减小了。这正如所期望的那样,因为根据式(3.4.3)和式(3.4.4),随着 K 的增加,均值图像越来越趋近于原始图像。我们还可以看到图 3.31 左列中差值图像减小均值的效果,K 越大则图像越暗。



(a) (b)

图 3.31 (a) 从上到下:图 3.30(a)分别与图 3.30(c)到(f)4幅图像间的差值图像,(b)相应的直方图

加处理是连续积分的离散形式。在天文观测中，与上述方法等效的一种方法是使用 CCD 或类似的传感器积累功能，通过长时间观测同一场景来减少噪声。然而，净效果与刚才讨论的相类似。我们可以通过冷却传感元件来进一步减少噪声。

正如在图像减处理过程中那样，当在 8 比特显示器上显示处理结果时，两幅或多幅 8 比特图像相加需要特别注意。由于 K 幅 8 比特图像求和的值的可能取值范围为 0 到 $255 \times K$，在这种情况下，简单地用 K 去除求和的结果把其重新标定为 8 比特，一般来说，在处理过程中精度会下降，但如果显示的范围被限制在 8 比特内，精度则不变。

当噪声加入一幅图像中时，图像求平均处理的某些实现中，有可能会出现负值。实际上，在刚刚给出的例子中，这是更精确的情况，因为，具有零均值和非零方差的高斯随机变量具有负值及正值。对上例中的图像我们采用了前节末尾讨论过的第二种标定方法。即，在给定的平均图像中，得到的最小值及它的负值加到图像上，然后，在改进后的图像中，采用 255/Max 去乘被改进图像的每一个像素的方法重新标定回[0,255]的范围，这里，Max 是该图像中的最大值。

3.5 空间滤波基础

在 3.1 节中已经提到，某些邻域处理工作是操作邻域的图像像素值以及相应的与邻域有相同维数的子图像的值。这些子图像可以被称为滤波器、掩模、核、模板或窗口，其中前三个词是更为普遍的术语。在滤波器子图像中的值是系数值，而不是像素值。

滤波的概念来源于在频率域对信号进行处理的傅里叶变换，这个题目将在第 4 章中详细阐述。在本章中，我们的兴趣在于直接对图像像素处理的操作。这里使用空间滤波这一词汇区别更传统的频率域滤波处理。

空间滤波的机理示于图 3.32。该处理就是在待处理图像中逐点地移动掩模。在每一点 (x,y) 处，滤波器在该点的响应通过事先定义的关系来计算。对于线性空间滤波（见 2.6 节线性的定义），其响应由滤波器系数与滤波掩模扫过区域的相应像素值的乘积之和给出。图 3.32所示为 3×3 的掩模，在图像中的点 (x,y) 处，用该掩模线性滤波的响应 R 为：

$$
\begin{aligned}
R = {} & w(-1,-1)f(x-1,y-1) + w(-1,0)f(x-1,y) + \cdots \\
& + w(0,0)f(x,y) + \cdots + w(1,0)f(x+1,y) + w(1,1)f(x+1,y+1)
\end{aligned}
$$

我们看到的是掩模系数与直接在掩模下的相应像素的乘积之和。注意，具体地说，系数 $w(0,0)$ 与图像值 $f(x,y)$ 相符合是指当乘积求和计算发生时，掩模位于 (x,y) 的中心。对于一个尺寸为 $m\times n$ 的掩模，我们假设 $m=2a+1$ 且 $n=2b+1$，这里 a,b 为非负整数。即，在后续的讨论中，处理的掩模的长与宽都为奇数，其有意义的最小尺寸为 3×3（在我们的讨论中排除了无足轻重的 1×1 掩模）。

一般来说，在 $M\times N$ 的图像 f 上，用 $m\times n$ 大小的滤波器掩模进行线性滤波由下式给出：

$$
g(x,y) = \sum_{s=-a}^{a}\sum_{t=-b}^{b} w(s,t)f(x+s,y+t) \tag{3.5.1}
$$

这里，与上一段一样 $a=(m-1)/2$ 且 $b=(n-1)/2$。为了得到一幅完整的经过滤波处理的图像，必须对 $x=0,1,2,\cdots,M-1$ 和 $y=0,1,2,\cdots,N-1$ 依次应用公式。这样，就保证了对图像

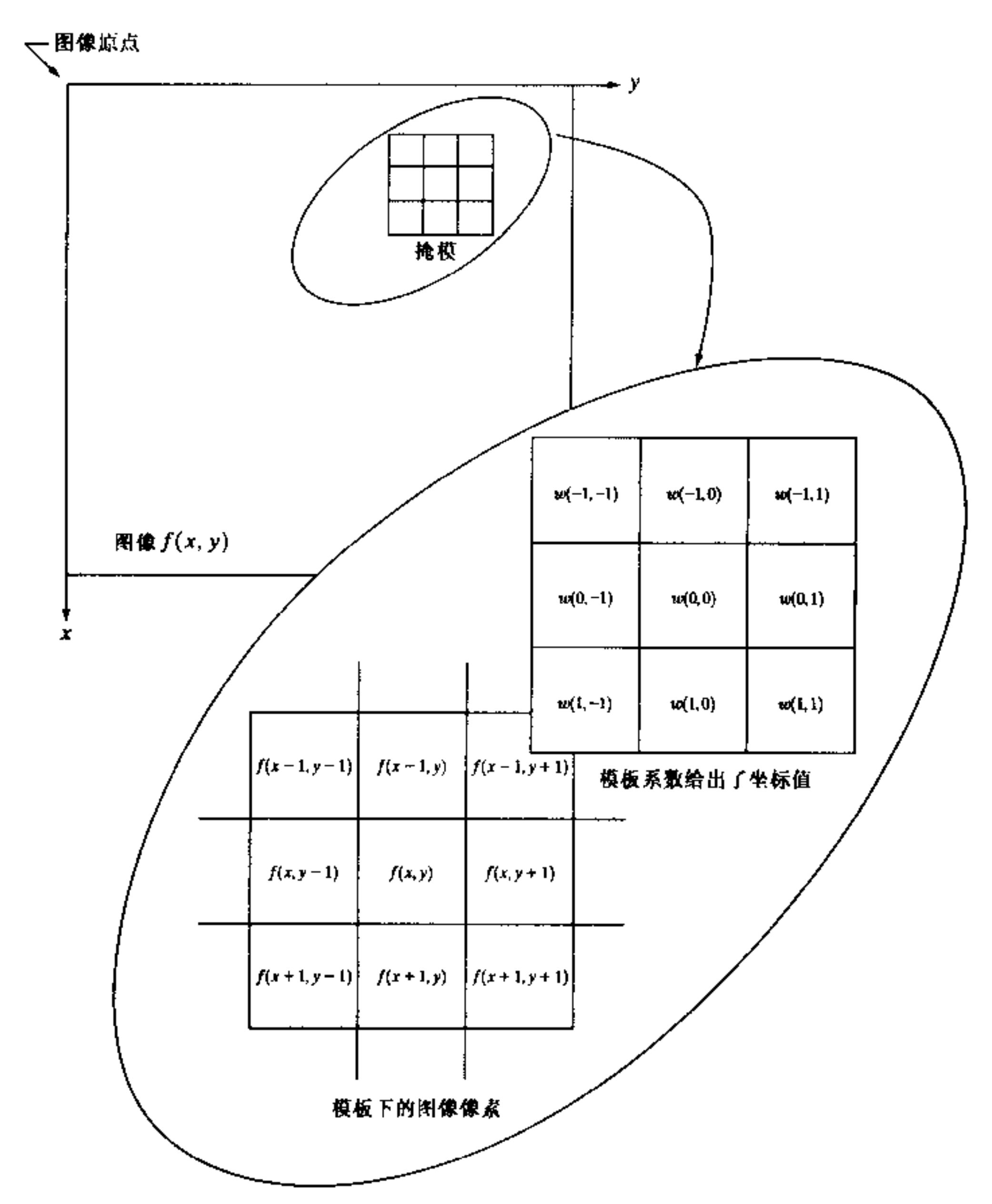


图 3.32 空间滤波的机理。放大的图显示了一个 3×3 掩模和其覆盖的图像部分。图像部分从掩模下移出，以便于观看

中的所有像素进行了处理。显然，当 $m = n = 3$ 时，就是上一段中给出例子的简化表达。

在第 4 章的讨论中，可以看到式(3.5.1)中的线性滤波处理与频率域中卷积处理的概念很相似，因此，线性空间滤波处理经常被称为"掩模与图像的卷积"。类似地，滤波掩模有时也可以称为"卷积模板"，"卷积核"一词也常用于此。

当我们的兴趣在于对图像任一点(x, y)进行 $m \times n$ 掩模处理得到的响应R，而不是模板卷积的机理时，实践中通常用如下形式简化表达形式：

$$R = w_1 z_1 + w_2 z_2 + \cdots + w_{mn} z_{mn} = \sum_{i=1}^{mn} w_i z_i \tag{3.5.2}$$

其中 W 为掩模系数，z 为与该系数对应的灰度值，mn 为掩模中包含的像素点总数。对于

图 3.33所示的 3×3 掩模,图像中任意一点(x,y)的响应由下式给出:

$$R = w_1 z_1 + w_2 z_2 + \cdots w_9 z_9 = \sum_{i=1}^{9} w_i z_i \tag{3.5.3}$$

因为在图像处理方面的文献中会经常看到该公式,所以对这一简单的公式应给予特别注意。

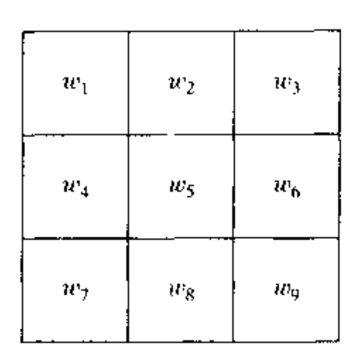


图 3.33　一般 3×3 空间滤波掩模的另一种表示

非线性空间滤波处理也是基于邻域处理,且掩模滑过一幅图像的机理与刚刚论述的一样。然而,一般说来,滤波处理取决于所考虑的邻域像素点的值,而不能直接用式(3.5.1)和式(3.5.2)中所描述的乘积求和中的系数。正如 3.6.2 节所示,利用非线性滤波器可以有效地降低噪声,这种非线性滤波器的基本函数是计算滤波器所在邻域的灰度中值。中值计算是非线性操作,正如在 3.3.4 节所使用的方差计算那样。

实现空间滤波邻域处理时的一个重要考虑因素就是,当滤波中心靠近图像轮廓时发生的情况。考虑一个简单的大小为 $n \times n$ 的方形掩模,当掩模中心距离图像边缘为$(n-1)/2$ 个像素时,该掩模至少有一条边与图像轮廓相重合。如果掩模的中心继续向图像边缘靠近,那么掩模的行或列就会处于图像平面之外。有很多方法可以处理这种问题。最简单的方法就是将掩模中心点的移动范围限制在距离图像边缘不小于$(n-1)/2$ 个像素处。这种做法将使处理后的图像比原始图像稍小,但滤波后的图像中的所有像素点都由整个掩模处理。如果要求处理后的输出图像与原始图像一样大,那么所采用的典型方法是,用全部包含于图像中的掩模部分滤波所有像素。通过这种方法,图像靠近边缘部分的像素带将用部分滤波掩模来处理。另一种方法就是在图像边缘以外再补上一行和一列灰度为零的像素点(其灰度也可以为其他常值),或者将边缘复制补在图像之外。补上的那部分经过处理后去除。这种方法保持了处理后的图像与原始图像尺寸大小相等,但是补在靠近图像边缘的部分会带来不良影响,这种影响随着掩模尺寸的增加而增大。总之,获得最佳滤波效果的惟一方法是使滤波掩模中心距原图像边缘的距离不小于$(n-1)/2$ 个像素。

3.6　平滑空间滤波器

平滑滤波器用于模糊处理和减小噪声。模糊处理经常用于预处理,例如,在提取大的目标之前去除图像中一些琐碎的细节、桥接直线或曲线的缝隙。通过线性滤波器和非线性滤波器的模糊处理可以减小噪声。

3.6.1 平滑线性滤波器

平滑线性空间滤波器的输出(响应)是包含在滤波掩模邻域内像素的简单平均值。因此,这些滤波器也称为均值滤波器。根据第4章的讨论,它们指的是低通滤波器。

平滑滤波器的概念非常直观。它用滤波掩模确定的邻域内像素的平均灰度值去代替图像每个像素点的值,这种处理减小了图像灰度的"尖锐"变化。由于典型的随机噪声由灰度级的尖锐变化组成,因此,常见的平滑处理应用就是减噪。然而,由于图像边缘(几乎总是一幅图像希望有的特性)也是由图像灰度尖锐变化带来的特性,所以均值滤波处理还是存在着不希望的边缘模糊的负面效应。另外,这类处理方法还有一些其他应用,比如2.4.3节讨论过的由于灰度量级不足而引起的伪轮廓效应的平滑处理。均值滤波器的主要应用是去除图像中的不相干细节,其中"不相干"是指与滤波掩模尺寸相比,较小的像素区域。

图3.34显示了两个3×3的平滑滤波器。第一个滤波器产生掩模下标准的像素平均值,这从把掩模系数代入式(3.5.3)可清楚地看出:

$$R = \frac{1}{9}\sum_{i=1}^{9} z_i$$

R是由掩模定义的3×3邻域像素灰度的平均值。注意,代替上式中的1/9,滤波器的系数全为"1"。这里的概念是,系数取1值时计算更有效。在滤波处理之后,整个图像被9除。一个$m \times n$掩模应有等于$1/mn$的归一化常数。所有系数都相等的空间均值滤波器有时称为盒滤波器。

图3.34中所示的第二种掩模更为重要一些。这个掩模也叫做加权平均,使用这一术语是指用不同的系数乘以像素,这样,从权值上看,一些像素比另一些更为重要。对于图3.34(b)所示的3×3掩模,处于掩模中心位置的像素比其他任何像素的权值都要大,因此,在均值计算中给定的这一像素显得更为重要。而距离掩模中心较远的其他像素就显得不太重要。由于对角项离中心比正交方向相邻的像素(参数为$\sqrt{2}$)更远,所以,它的重要性要比与中心直接相邻的四个像素低。把中心点加强为最高,而随着距中心点距离的增加减小系数值,是为了减小平滑处理中的模糊。我们也可以采取其他权重达到相同的目的。然而,图3.34(b)掩模中的所有系数的和应该为16,这很便于计算机的实现,因为它是2的整数次幂。在实践中,由于这些掩模在一幅图像中所占的区域很小,通常很难看出使用图3.34的各种掩模或用其他类似手段平滑处理后的图像之间的区别。

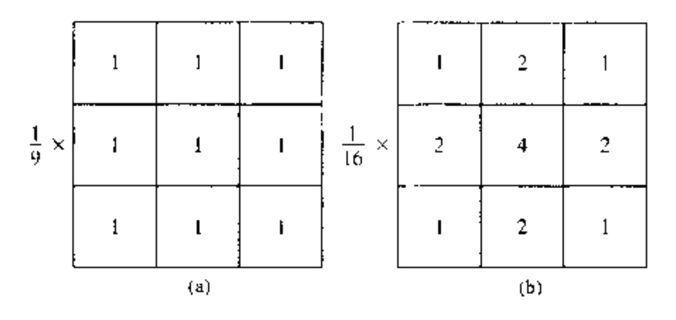


图3.34 两个3×3平滑(均值)滤波器掩模。每一个掩模前边的乘数等于它的系数值的和,以计算平均值

参见式(3.5.1),一幅 $M\times N$ 的图像经过一个 $m\times n$(m 和 n 是奇数)的加权均值滤波器滤波的过程可由下式给出:

$$g(x,y)=\frac{\sum_{s=-a}^{a}\sum_{t=-b}^{b}w(s,t)f(x+s,y+t)}{\sum_{s=-a}^{a}\sum_{t=-b}^{b}w(s,t)} \tag{3.6.1}$$

这一式中的参数见式(3.5.1)的定义。正如前面说过的,它可以这样理解,即一幅完全滤波的图像是由对 $x=0,1,2,\cdots,M-1$ 与 $y=0,1,2,\cdots,N-1$ 执行式(3.6.1)得到的。式(3.6.1)中的分母部分简单地表示为掩模的系数总和,而且因为它是一个常数,只需计算一次就可以了。这个比例参数在滤波处理完成之后一般用于输出图像的所有像素。

例 3.9　用各种尺寸的掩模平滑图像

与滤波掩模大小有关的图像平滑效果如图 3.35 所示,图中显示了一幅原始图像以及分别用尺寸为 $n=3,5,9,15$ 和 35 像素的方形均值滤波器得到的相应平滑结果。这些结果的主要特点有:当 $n=3$ 时,可以观察到在整幅图像中有轻微的模糊,但是,正如所希望的那样,当图像细节与滤波器掩模近似相同时,图像中一些细节受到的影响比较大。例如,用 3×3 和 5×5 矩形掩模时,较小的字母"a"和细的颗粒要比图像其他部分更为模糊。一个正面效果是噪声明显地减少了。注意,字母的锯齿状边界与圆圈都被平滑了。

对于 $n=5$ 的结果基本类似,模糊程度稍微有所增加。对于 $n=9$,明显地看出更为模糊了,而且有 20% 的黑色圆圈几乎不能像前三幅图像那样与背景明显区分开来。这些说明,当目标与其相邻像素的灰度值接近时,混合效应会导致目标模糊。当 $n=15$ 和 35 时,就图像中目标物的尺寸来说,结果已是极端情况了。我们经常运用这些极端类型的模糊处理来去除图像中的一些小物体。例如,在图 3.35(f)中,三个小方框、两个小圆圈以及大部分矩形的噪声区域已经融入到背景中了。注意,在这张图中,还有明显的边界现象。这是我们将 0 添加到原始图像的边缘上,经过处理后再去除添加区域的结果。某些黑色混入了滤波后的图像,但是,对于用较大滤波器平滑的图像,这就成问题了。

前面曾经提到,空间均值处理的重要应用是,为了对感兴趣物体得到一个粗略的描述而模糊一幅图像,这样,那些较小物体的强度与背景混合在一起了,较大物体变得像"斑点"而易于检测。掩模的大小由那些即将融入背景中去的物件尺寸来决定。作为实例,考虑图 3.36(a),它是绕地球轨道上的 Hubble 望远镜拍摄下来的一幅图像。图 3.36(b)显示了应用 15×15 的均值滤波器掩模对该图像处理后的结果。可以看到,图片中的一些部分或者融入到背景之中,或者其亮度明显地减小了。像这样利用阈值并基于物体亮度来消除某些目标的操作是很典型的。用等于模糊图像最高亮度 25% 的阈值,并使用图 3.2(b)的阈值函数处理的结果示于图 3.36(c)。用原图像比较经过处理的结果,我们可以看到考虑图像中最大、最亮目标的表达方式是非常合理的。

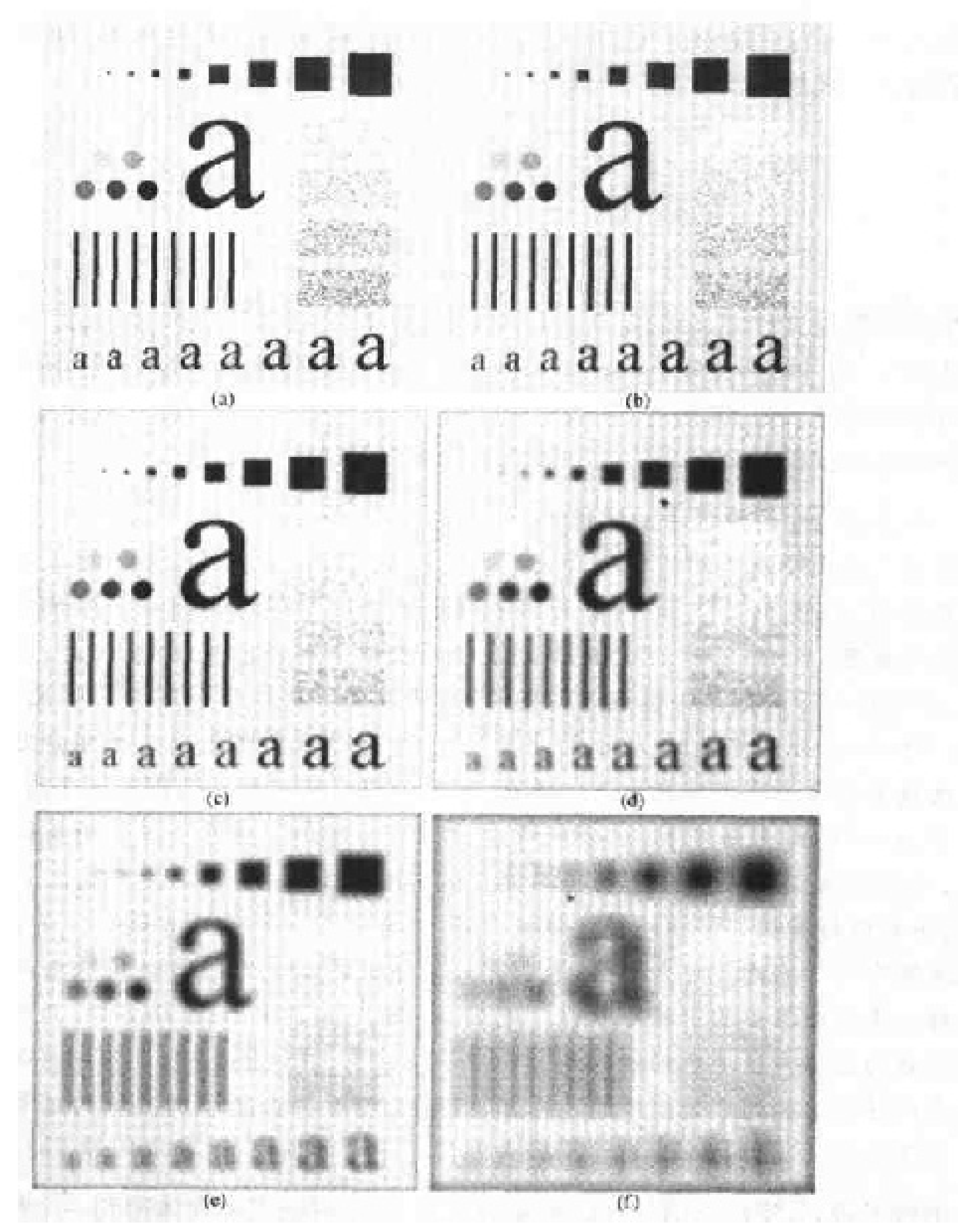
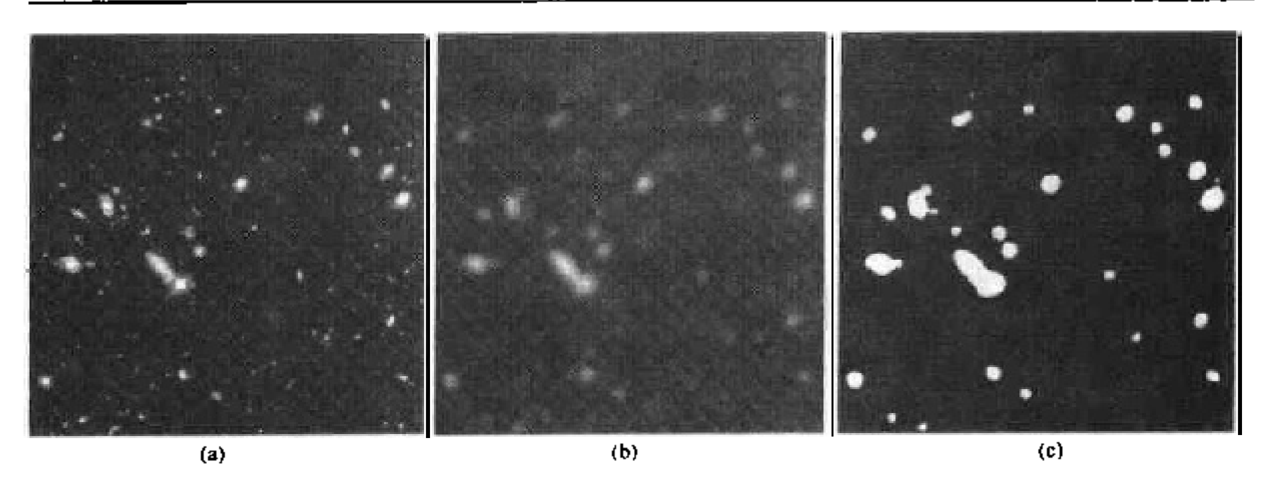


图3.35 (a)大小为500×500像素的原图像,(b)到(f)用大小为 $n=3,5,9,15$ 和35的方形均值滤波掩模平滑处理的结果。在顶端的黑方块大小分别为3,5,9,15,25,35,45和55个像素,边界相隔25个像素,位于底端的字母大小在10到24点之间,增量为2个点;顶端最大的字母是60个点。垂直线段是5个像素宽,100个像素高;它们的间隔是20个像素。圆的直径是25个像素,它们的边界相隔15个像素,灰度级在0%到100%范围内,从黑以20%增加。图像的背景是黑色的10%。噪声矩形大小是50×120像素



(a)　(b)　(c)

图 3.36　(a)取自 Hubble 空间望远镜的图像,(b)由 15×15 均值掩模处理过的图像,(c)对图(b)进行阈值处理的结果(原图像由NASA提供)

3.6.2　统计排序滤波器

统计滤波器是一种非线性的空间滤波器,它的响应基于图像滤波器包围的图像区域中像素的排序,然后由统计排序结果决定的值代替中心像素的值。统计滤波器中最常见的例子就是中值滤波器,正如其名,它是将像素(在中值计算中包括的原像素值)邻域内灰度的中值代替该像素的值。中值滤波器的使用非常普遍,这是因为对于一定类型的随机噪声,它提供了一种优秀的去噪能力,比小尺寸的线性平滑滤波器的模糊程度明显要低。中值滤波器对处理脉冲噪声(也称为椒盐噪声)非常有效,因为这种噪声是以黑白点叠加在图像上的。

一个数值集合的中值 ξ 是这样的数值,即,数值集合中,有一半小于或等于 ξ,还有一半大于或等于 ξ。为了对一幅图像上的某个点做中值滤波处理,必须先将掩模内欲求的像素及其邻域的像素值排序,确定出中值,并将中值赋予该像素点。例如,对于一个 3×3 的邻域,其中值是第 5 个值,而在一个 5×5 的邻域中,中值就是第 13 个值,等等。当一个邻域中的一些像素值相同时,它们中的任何一个都可以作为中值。例如,在一个 3×3 的邻域内有一系列像素值(10,20,20,20,15,20,20,25,100),对这些值排序后为(10,15,20,20,20,20,20,25,100),那么其中值就是 20。这样,中值滤波器的主要功能是使拥有不同灰度的点看起来更接近于它的邻近值。事实上,是用 $n \times n$ 的中值滤波器去除那些相对于其邻域像素更亮或更暗,并且其区域小于 $n^2/2$(滤波器区域的一半)的孤立像素集。在这种情况下,"去除"的意思是强制为邻域的中间亮度。而对较大的像素集的影响明显减小。

尽管在图像处理中,中值滤波器是用得最广泛的统计滤波器,但这并不等于它是惟一的。所谓中值就是一系列像素值的第50% 个值,但读者根据基本统计学知道,排序也适用于其他不同的情况。例如,我们可以取第100% 个值,即我们所说的最大值滤波器,这种滤波器在搜寻一幅图像中的最亮点时非常有用。一个 3×3 的最大值滤波器的响应可以用公式 $R = \max\{z_k \mid k = 1,2,3,\cdots,9\}$ 来给出。相反,当取 0% 的值时就变成了最小值滤波器,它可用于相反的目的。中值、最大值和均值滤波器将在第 5 章中做详细介绍。

例 3.10　利用中值滤波器降噪

图 3.37(a)显示了一张受椒盐噪声污染的电路板的 X 光图像。为了能够说明这种情况

下,中值滤波器的处理效果比均值滤波器更好,在图 3.37(b)中,我们展示了用 3×3 均值滤波掩模处理噪声图像的结果,而图 3.37(c)则为经过 3×3 中值滤波器处理后的结果。经均值滤波器处理后的图像中虽然可见噪声更少很明显,但所付出的代价是图像变得模糊了。在这种情况下综合考虑可以得出,中值滤波器要远远优于均值滤波器。通常,中值滤波器比均值滤波器更适合去除加性椒盐噪声。

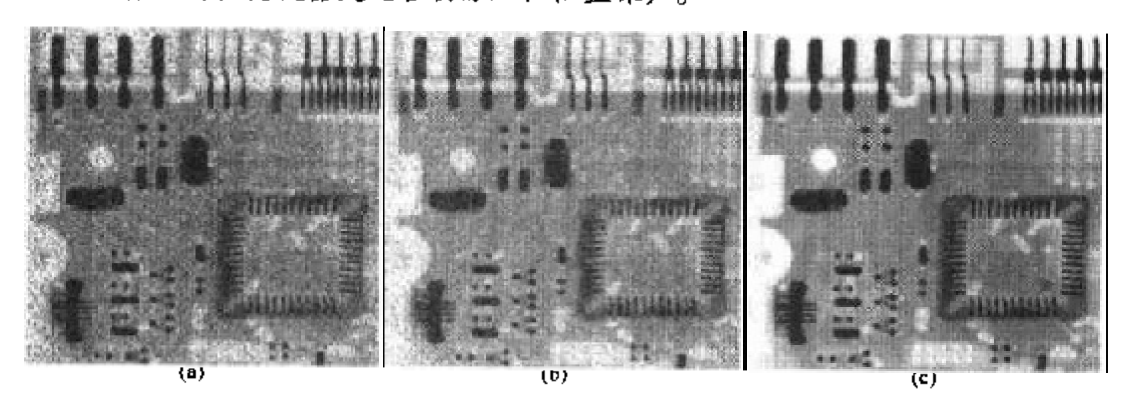


图 3.37 (a)椒盐噪声污染的电路板 X 光图像,(b)用 3×3 均值掩模去除噪声,(c)用3×3中值滤波器去除噪声

3.7 锐化空间滤波器

锐化处理的主要目的是突出图像中的细节或者增强被模糊了的细节,这种模糊不是由于错误操作,就是特殊图像获取方法的固有影响。图像锐化处理的方法多种多样,其也包括多种应用,从电子印像和医学成像到工业检测和军事系统的制导,等等。

在最后一节中,我们将看到在空间域用像素邻域平均法可以使图像变模糊。因为均值处理与积分相类似,从逻辑角度我们可以断定,锐化处理可以用空间微分来完成。在这一节中将讨论数字微分锐化的各种定义及其实现算子。总的来说,微分算子的响应强度与图像在该点(应用了算子)的突变程度有关。这样一来,图像微分增强了边缘和其他突变(如噪声)并削弱了灰度变化缓慢的区域。

3.7.1 基础

在以下两节中,我们将分别讨论基于一阶和二阶微分的细节锐化滤波器。在讨论具体滤波器之前,还是先回顾一下数学中微分的某些基本性质。为了说明简单,主要集中讨论一阶微分的性质。我们最感兴趣的微分性质是恒定灰度区域(平坦段)、突变的开头与结尾(阶梯和斜坡突变)及沿着灰度级斜坡处的特性。这些类型的突变可以用来对图像中的噪声点、细线与边缘模型化。在向(从)这些图像特性过渡时的微分性质也很重要。

数学函数的微分可以用不同的术语定义,也有各种方法定义这些差别,然而,对于一阶微分的任何定义都必须保证以下几点:(1)在平坦段(灰度不变的区域)微分值为零;(2)在灰度阶梯或斜坡的起始点处微分值非零;(3)沿着斜坡面微分值非零。任何二阶微分的定义也类似:(1)在平坦区微分值必为零;(2)在灰度阶梯或斜坡的起始点处微分值非零;(3)沿着斜坡面微分值非零。因为我们处理的是数字量,其值是有限的,故最大灰度级的变化也是有限的,变化

发生的最短距离是在两相邻像素之间。对于一元函数 $f(x)$ 表达一阶微分的定义是一个差值：

$$\frac{\partial f}{\partial x} = f(x+1) - f(x)$$

这里，为了与对二元图像函数 $f(x,y)$ 求微分时的表达式保持一致，使用偏导数符号。对二元函数，我们将沿着两个空间轴处理偏微分。当前讨论的空间微分的应用并不影响我们试图完成的任何方法的本质。

类似地，用差分定义二阶微分：

$$\frac{\partial^2 f}{\partial x^2} = f(x+1) + f(x-1) - 2f(x)$$

很容易证实这两个定义满足前面所说的一阶、二阶微分的条件。为了解这一点，研究示于图 3.38的例子，并强调一下在图像处理中一阶和二阶微分间的相同及不同点。

图 3.38(a) 是一幅简单图像，其中包含各种实心物体、一条线及一个单一噪声点。图 3.38(b)是沿着中心并包含噪声点的此图像的水平剖面图。这张剖面图是将要用以说明该图的一维函数。图 3.38(c)示出的是简化的剖面图，在这张图中我们取了足够多的点以便于分析噪声点、线及物体边缘的一阶和二阶微分结果。在简化图中，斜坡的过渡包含四个像素，噪声点是一个单一像素，线有三个像素粗，而灰度阶梯的过渡变化在相邻像素之间发生。灰度级数目简化为只有 8 个等级。

从左向右横穿剖面图讨论一阶和二阶微分的性质。首先，我们注意到，沿着整个斜坡，一阶微分值都不是零，而经二阶微分后，非零值只出现在斜坡的起始处和终点处。因为在图像中，边缘类似这种类型的过渡，由此，我们得出结论，一阶微分产生较粗的边缘，而二阶微分则细得多。其次，我们来讨论孤立的噪声点。这里，在该噪声点及周围点上，二阶微分比一阶微分的响应要强很多，当然，这是我们所不希望的。在进行锐度变化增强的处理中，二阶微分比一阶微分更好，所以，可以预料在做细节增强处理时二阶微分比一阶微分强得多。细线可以看做一细节，基本可以看到两种微分处理后的同样的区别。如果这条细线的最大灰度值与孤立点相同，那么经二阶微分后的响应对于后者更强烈。最后，在本例中，灰度阶梯上的两种微分结果相同(在大部分情况下，不是从 0 过渡到阶梯时，二阶微分结果的灰度级更弱一些)。我们还注意到，二阶微分有一个过渡，即从正回到负。在一幅图像中，该现象表现为双线。这一双边缘效果在第 10 章中将作为一个重要问题讨论。这里，我们把微分用于边缘检测。另外，我们还注意到重要的一点，即如果细线的灰度与阶梯相同，那么对二阶微分处理的响应，细线要比阶梯强。

总之，通过比较一阶微分处理与二阶微分处理的响应，我们得出以下结论：(1)一阶微分处理通常会产生较宽的边缘；(2)二阶微分处理对细节有较强的响应，如细线和孤立点；(3)一阶微分处理一般对灰度阶梯有较强的响应；(4)二阶微分处理对灰度级阶梯变化产生双响应。我们还注意到，二阶微分在图像中灰度值变化相似时，对线的响应要比对阶梯强，且点比线响应强。

大多数应用中，对图像增强来说，二阶微分处理比一阶微分好一些，因为形成增强细节的能力好一些。由于这一原因及实现和扩展都简单，对图像增强我们开始注意应用二阶微分处理。一阶微分处理将在 3.7.3 节中讨论。尽管一阶微分在图像处理中主要用于边缘提取，但它们在图像增强中也起着很大作用。事实上，我们将在 3.8 节中与二阶微分结合起来应用以达到更好的增强效果。

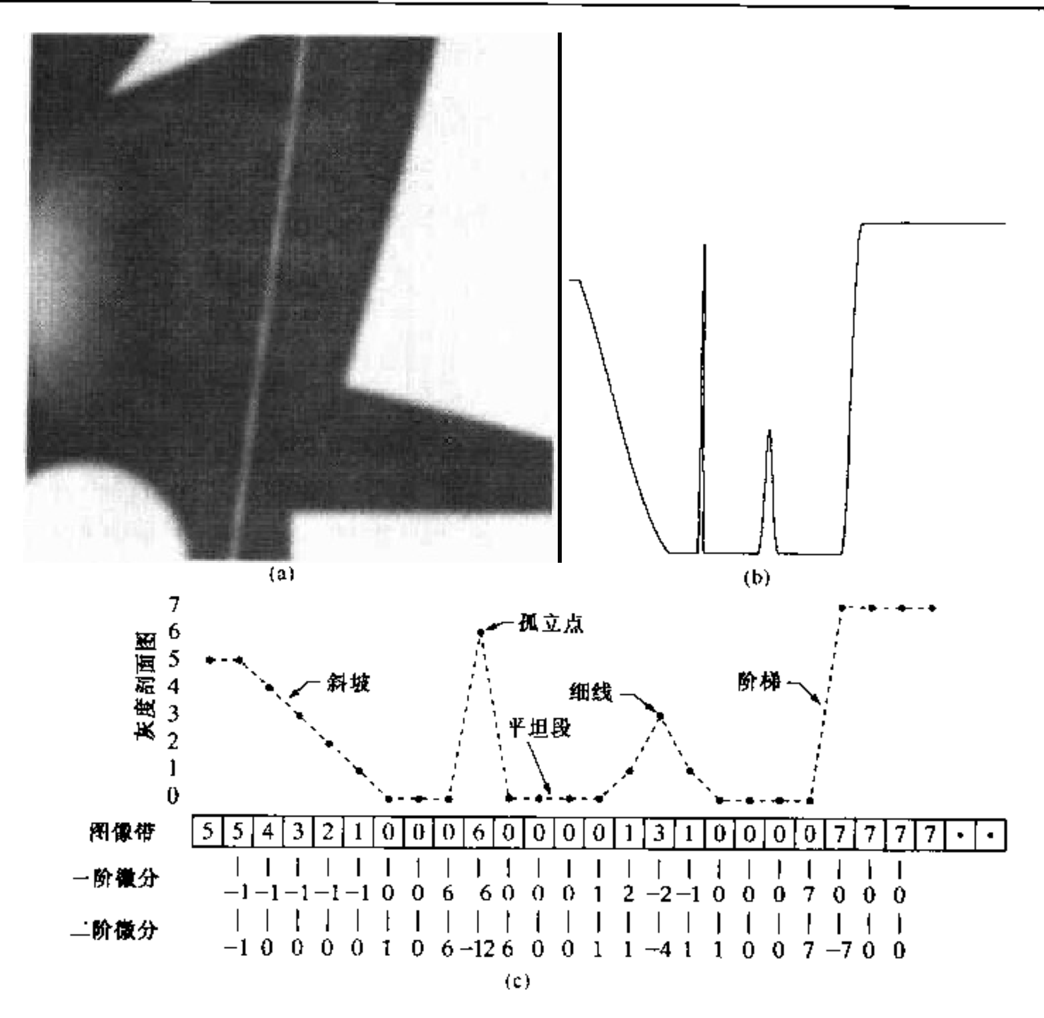


图 3.38 (a)一幅简单图像,(b)沿图像中心并且包含孤立噪声点的一维水平灰度剖面图,(c)简化的剖面图(用虚线连接点以简化描述)

3.7.2 基于二阶微分的图像增强——拉普拉斯算子

在本节中,将详细介绍二元函数的二阶微分在图像增强处理中的应用。首先定义一个二阶微分的离散公式,然后构造一个基于此式的滤波器。我们最关注的是一种各向同性滤波器,这种滤波器的响应与滤波器作用的图像的突变方向无关。也就是说,各向同性滤波器是旋转不变的,即将原始图像旋转后进行滤波处理给出的结果与先对图像滤波,然后再旋转的结果相同。

处理方法

可以看出(Rosenfeld 和 Kak[1982])最简单的各向同性微分算子是拉普拉斯算子,一个二元图像函数 $f(x,y)$ 的拉普拉斯变换定义为:

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{3.7.1}$$

因为任意阶微分都是线性操作,所以拉普拉斯变换也是一个线性操作。

为了更适合于数字图像处理,这一方程需要表示为离散形式。通过邻域处理有多种方法定义离散变换,但无论怎样定义,都必须符合 3.7.1 节中提到的二阶微分处理的性质。在上一

节中给出的数字二阶微分处理的定义是最常用的一种。考虑到有两个变量，因此，我们在 x 方向上对二阶偏微分采用下列定义：

$$\frac{\partial^2 f}{\partial^2 x^2} = f(x+1,y) + f(x-1,y) - 2f(x,y) \tag{3.7.2}$$

类似地，在 y 方向上为：

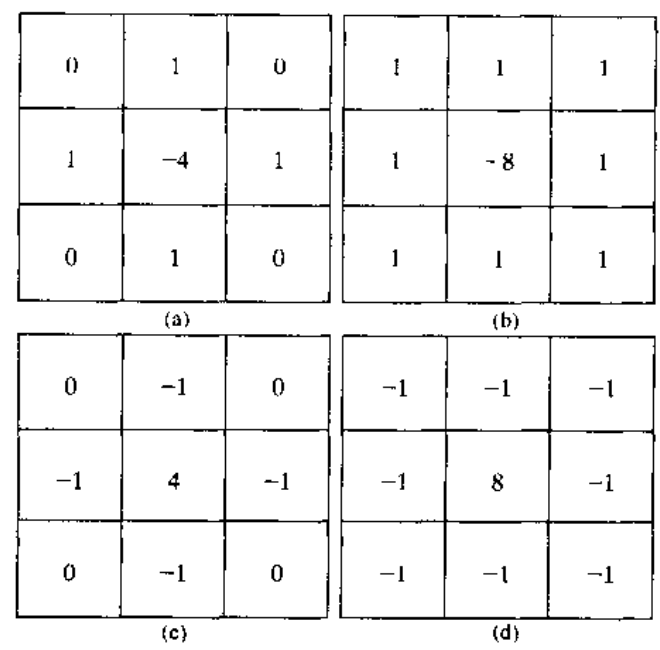
$$\frac{\partial^2 f}{\partial^2 y^2} = f(x,y+1) + f(x,y-1) - 2f(x,y) \tag{3.7.3}$$

式(3.7.1)中的二维拉普拉斯数字实现可由这两个分量相加得到：

$$\nabla^2 f = [f(x+1,y) + f(x-1,y) + f(x,y+1) + f(x,y-1)] - 4f(x,y) \tag{3.7.4}$$

这个公式可以用图 3.39(a)所示的掩模来实现，它们给出了以 90°旋转的各向同性的结果。实现机理在式(3.5.1)中给出，并在 3.6.1 节线性平滑滤波器中已讲述过，我们在这里只是简单地使用了不同的系数。

对角线方向也可以加入到离散拉普拉斯变换的定义中，只需在式(3.7.4)中添入两项，即两个对角线方向各加一个。每一个新添加项的形式与式(3.7.2)或式(3.7.3)类似，只是其坐标轴的方向沿着对角线方向。由于每个对角线方向上的项还包含一个 $-2f(x,y)$，所以，现在从不同方向的项上减去的总和是 $-8f(x,y)$。执行这一新定义的掩模如图 3.39(b)所示。这种掩模对 45°增幅的结果是各向同性的。图 3.39 所示的另外两个掩模在实践中也经常使用。这两个掩模也是以拉普拉斯变换定义为基础的，只是其中的系数与我们在这里所用到的符号相反而已。正因如此，它们产生等效的结果，但是，当拉普拉斯滤波后的图像与其他图像合并时(相加或相减)，则必须考虑符号上的差别。



(a)

0	1	0
1	−4	1
0	1	0

(b)

1	1	1
1	−8	1
1	1	1

(c)

0	−1	0
−1	4	−1
0	−1	0

(d)

−1	−1	−1
−1	8	−1
−1	−1	−1

图 3.39　(a)执行式(3.7.4)定义的离散拉普拉斯变换所用的滤波器掩模，(b)用于执行该公式的扩展掩模，它包括对角线邻域，(c)和(d)其他两种拉普拉斯的实现

由于拉普拉斯是一种微分算子，它的应用强调图像中灰度的突变及降低灰度慢变化的区域。这将产生一幅把图像中的浅灰色边线和突变点叠加到暗背景中的图像。将原始图像和拉

普拉斯图像叠加在一起的简单方法可以保护拉普拉斯锐化处理的效果,同时又能复原背景信息。正如上一段中讲到的,记住所使用的拉普拉斯定义是很重要的。如果所使用的定义具有负的中心系数,那么,就必须将原始图像减去经拉普拉斯变换后的图像而不是加上它,从而得到锐化的结果。所以,我们使用拉普拉斯变换对图像增强的基本方法可表示为下式:

$$g(x,y)=\begin{cases} f(x,y)-\nabla^2 f(x,y) & \text{如果拉普拉斯掩模中心系数为负} \\ f(x,y)+\nabla^2 f(x,y) & \text{如果拉普拉斯掩模中心系数为正} \end{cases} \tag{3.7.5}$$

这一公式的应用在后面说明。

例 3.11 用拉普拉斯的图像锐化

图 3.40(a)显示了一幅月球北极的照片。图 3.40(b)显示了用图 3.39(b)中的拉普拉斯掩模对该图像滤波后的结果。由于拉普拉斯图像中既有正值又有负值,一个典型的标定方法是使用在 3.4.1 节中末尾讨论的方法。为了这一目的,有时会用到绝对值,但是,这的确是不正确的,因为,它会产生近似等于其大小的双线,使图像变模糊。

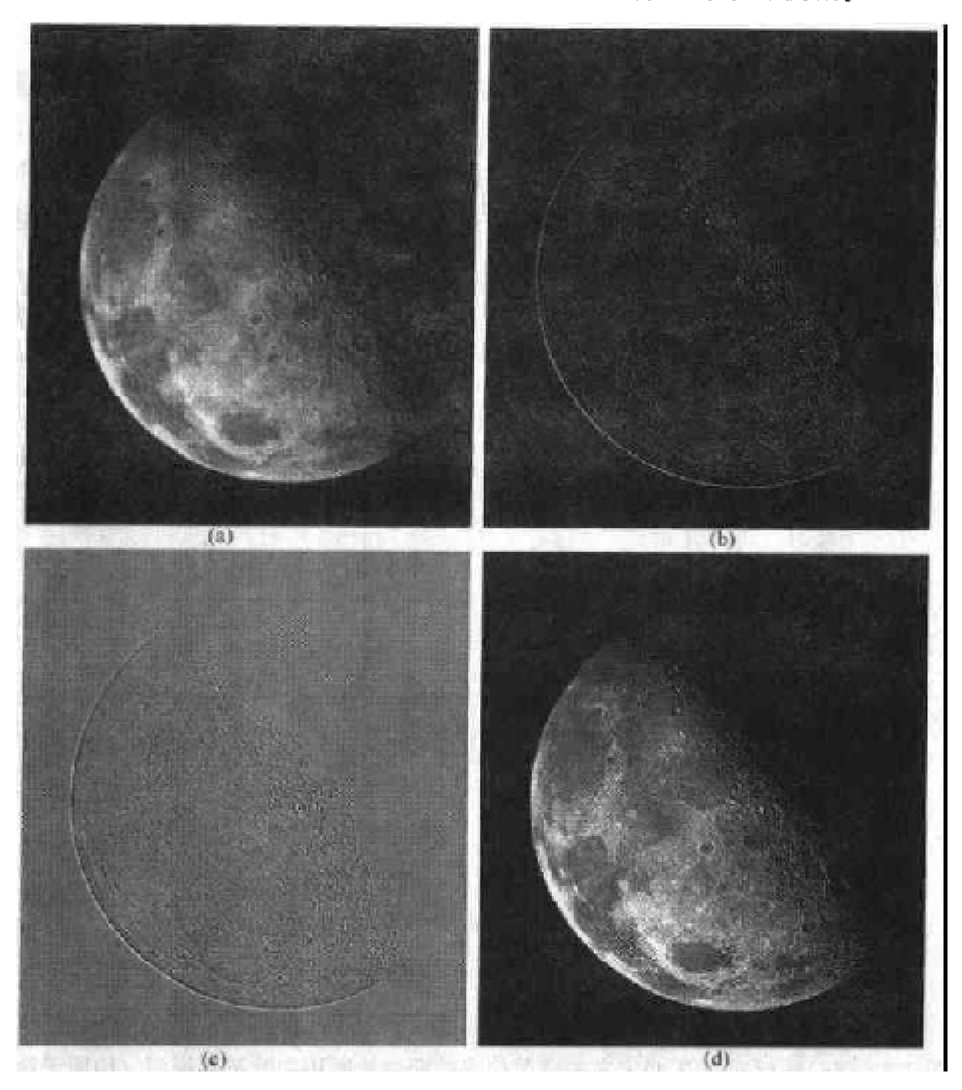


图 3.40 (a)月球北极的图像,(b)拉普拉斯滤波后的图像,(c)为显示目的标定后的拉普拉斯图像,(d)用式(3.7.5)增强后的图像(原图像由 NASA 提供)

为了显示,图3.40(c)所示的图像用刚才描述的方法进行了标定。注意,这幅图像最突出的特点是边缘及各级灰度值突变处的锐化灰度级。前边谈到的接近黑色的背景由于标定而呈灰色。这一呈现浅灰色的现象是典型的被适当标定的拉普拉斯图像。最后,图3.40(d)显示了用式(3.7.5)处理得到的结果。该图像的细节比原始图像更加清晰。将原始图像加到拉普拉斯的处理结果中去就可以使图像中的各灰度值得到复原,而且通过拉普拉斯变换增强了图像中灰度突变处的对比度。最终结果是使图像中小的细节部分得到增强并良好保留了图像的背景色调。基于拉普拉斯变换的图像增强已成为图像锐化处理的一个基本工具。

简化

在上一例中,我们首先计算拉普拉斯过滤图像,然后,从原始图像减掉该图像来实现式(3.7.5)。这样做是为了教学目的,以说明处理过程中每一步的作用。在实际运用中,式(3.7.5)通常用单一掩模的一次扫描来实现。单一掩模的系数很容易由式(3.7.5)中的第一行替代式(3.7.4)中的$\nabla^2 f(x,y)$得到:

$$\begin{aligned} g(x,y) &= f(x,y) - [f(x+1,y) + f(x-1,y) \\ &\quad + f(x,y+1) + f(x,y-1)] + 4f(x,y) \\ &= 5f(x,y) - [f(x+1,y) + f(x-1,y) \\ &\quad + f(x,y+1) + f(x,y-1)] \end{aligned} \tag{3.7.6}$$

该式通过图3.41(a)所示的掩模来实现。如果对角线方向上的邻值也包含在拉普拉斯计算中,应该使用图3.41(b)中所示的掩模。如果用式(3.7.5)中的第二行代替式(3.7.4)的负值,将会产生相同的掩模。

例3.12 使用复合拉普拉斯掩模的图像增强

经包含有对角分量的掩模进行处理得到的结果一般比用图3.41(a)那些更基本掩模处理的所得结果更为锐化。这种性质可以用图3.41(d)和(e)所示的经拉普拉斯滤波处理的图像加以说明,其中图(d)与(e)分别是原始图像通过图3.41(a)与(b)中的掩模处理的结果。通过将被处理的图像与原始图像3.41(c)进行比较,可以看出这两种掩模产生了有效的增强,但是,使用图3.41(b)掩模处理后的结果锐化得更明显。图3.41(c)是一幅烧坏的钨丝的扫描电子显微镜图像,放大了250倍。

由于拉普拉斯变换是线性算子,可以将式(3.7.5)看成两种线性处理之差,从而得到与图3.41(a)和图3.41(b)中相同的混合掩模。即,把$f(x,y)$看成是用一个掩模处理其自身,该掩模在中心处有一个单位系数而其他地方系数均为0。方程的第二项则是一幅用图3.39的拉普拉斯掩模之一处理过的相同图像。由于是线性,在式(3.7.5)中用单位中心掩模及那些拉普拉斯掩模之一得到的结果将与从单位中心掩模中加上或减去拉普拉斯掩模形成的单一掩模得到的结果相同。

反锐化掩蔽与高提升滤波处理

长期以来在出版业中使用的图像锐化处理是将图像模糊形式从原始图像中去除。这种处理被称为图像的反锐化掩蔽,可以表示为:

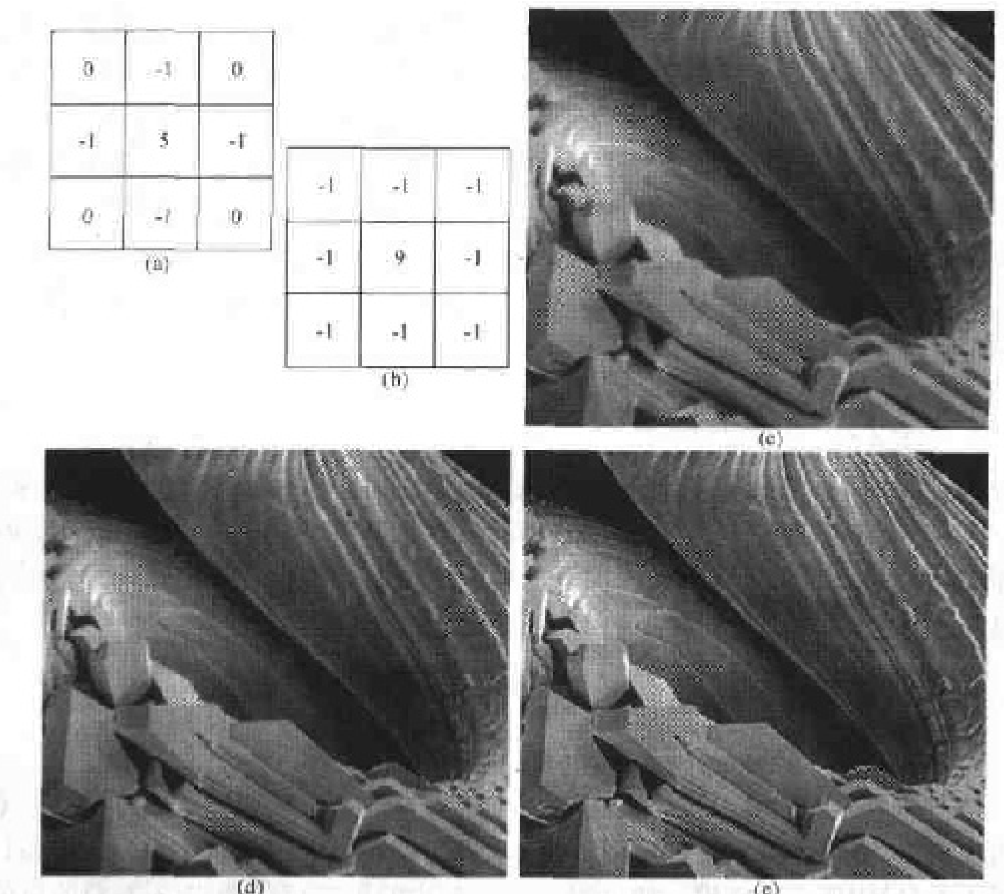


图 3.41 (a)合成拉普拉斯掩模,(b)第二种合成掩模,(c)扫描电子显微镜图像,(d)和(e)分别用(a)和(b)掩模滤波的结果,注意,(e)比(d)锐化得多(原图像由俄勒冈大学地质系的Michael Shaffer先生提供)

$$f_s(x,y) = f(x,y) - \bar{f}(x,y) \quad (3.7.7)$$

其中 $f_s(x,y)$ 表示经过反锐化掩蔽得到的锐化图像,是 $f(x,y)$ 的模糊形式。反锐化掩蔽处理最早应用于摄影暗室中,将一张模糊的负片与相应的正片卷合在一起,然后,冲洗这一混合的胶片得到一张更为清晰的照片。

反锐化掩蔽进一步的普遍形式称为高提升滤波。在图像中任何一点 (x,y) 处,高提升滤波后的图像 f_{hb} 可定义如下:

$$f_{hb}(x,y) = Af(x,y) - \bar{f}(x,y) \quad (3.7.8)$$

其中 $A \geqslant 1$,与前式一样,是 $\bar{f}$ 的模糊形式,此式也可以写成:

$$f_{hb}(x,y) = (A-1)f(x,y) + f(x,y) - \bar{f}(x,y) \quad (3.7.9)$$

结合式(3.7.7),我们可以得到:

$$f_{hb}(x,y) = (A-1)f(x,y) + f_s(x,y) \quad (3.7.10)$$

这一表达式可计算高提升滤波图像。

式(3.7.10)在一般情况下都成立,但并不能说明锐化图像是怎么得到的。如果我们选择拉普拉斯变换,就可以知道 $f_s(x,y)$ 是通过式(3.7.5)得到的。在这种情况下,式(3.7.10)变成:

$$f_{hb}=\begin{cases}Af(x,y)-\nabla^2 f(x,y) & \text{如果拉普拉斯掩模中心系数为负}\\ Af(x,y)+\nabla^2 f(x,y) & \text{如果拉普拉斯掩模中心系数为正}\end{cases}\tag{3.7.11}$$

高提升滤波处理可以通过任何一个图 3.42 所示的掩模得以实现。我们注意到，当 $A=1$ 时，高提升滤波处理就是标准的拉普拉斯变换。随着 A 超过 1 不断增大，锐化处理的效果越来越不明显。最终，当 A 足够大时，高提升图像将近似等于经常数调制的图像。

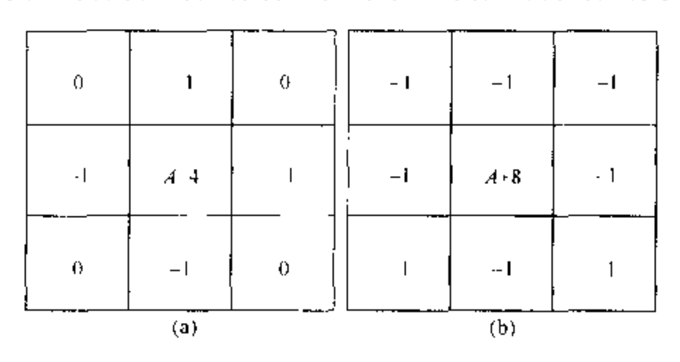


图 3.42　高提升滤波技术可以用其中一种掩模来实现（$A\geqslant 1$）

例 3.13　用高提升滤波器增强图像

提升滤波的主要应用之一是输入图像太暗时的处理。通过使用不同的提升系数，通常可以使图像整体的平均灰度值增加，从而使最后结果提高图像的亮度。图 3.43 向我们展示了这一应用。图 3.43(a)与图 3.41(c)相比较暗。图 3.43(b)是经过图 3.42(b)中所示的拉普拉斯掩模处理的结果，其中 $A=0$。图 3.43(c)是经过图 3.42(b)所示的拉普拉斯掩模处理的结果，其中 $A=1$。正如我们预料的那样，图像被锐化了，但结果仍然像原始图像那么暗。最后，图 3.43(d)显示了用 $A=1.7$ 处理的结果。这是一个更能接受的处理结果，其平均灰度增大了，因此图像看上去更亮、更自然了。

3.7.3　基于一阶微分的图像增强——梯度法

在图像处理中，一阶微分是通过梯度法来实现的。对于函数 $f(x,y)$，在其坐标 (x,y) 上的梯度是通过一个二维列向量来定义的：

$$\nabla \mathbf{f}=\begin{bmatrix}G_x\\ G_y\end{bmatrix}=\begin{bmatrix}\dfrac{\partial f}{\partial x}\\ \dfrac{\partial f}{\partial y}\end{bmatrix}\tag{3.7.12}$$

这个向量的模值由下式给出：

$$\begin{aligned}\nabla f&=\text{mag}(\nabla \mathbf{f})\\ &=\left[G_x^2+G_y^2\right]^{\frac{1}{2}}\\ &=\left[\left(\frac{\partial f}{\partial x}\right)^2+\left(\frac{\partial f}{\partial y}\right)^2\right]^{\frac{1}{2}}\end{aligned}\tag{3.7.13}$$

尽管梯度向量的分量本身是线性算子，但这一向量的模值显然不是线性的，这是由于用到了平方和开放运算。另外，式(3.7.12)中的偏导数并非是旋转不变的（各向同性），但梯度向量的模值却是各向同性的。尽管这样说在严格意义上并不正确，但我们一般把梯度矢量的模值称为

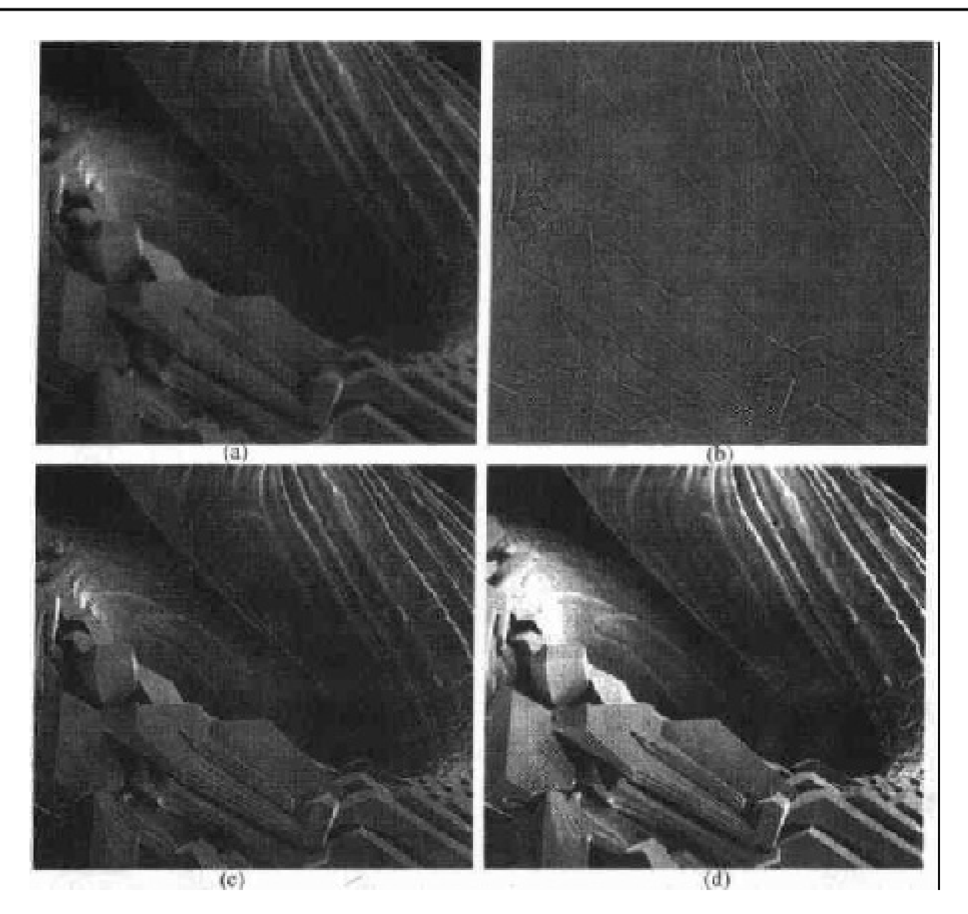


图 3.43 (a)与图 3.41(c)相同,但较暗,(b)用图 3.42(b)中所示的拉普拉斯掩模处理的结果,$A=0$,(c)使用图 3.42(b)所示的拉普拉斯掩模增强的图像,其中 $A=1$,(d)与(c)相同,但 $A=1.7$

梯度。为保持惯例,在以下的讨论中将使用这一术语,只有当它们两者会引起混淆时,才对向量和它的模值加以明确区分。

当对整幅图像进行式(3.7.13)的计算时运算量很大,因此,在实际操作中,常用绝对值代替平方与平方根运算近似求梯度的模值:

$$\nabla f \approx |G_x| + |G_y| \tag{3.7.14}$$

这个公式计算起来较为简单并且保持着灰度的相对变化,但各向同性特性通常就不存在了。然而,对于拉普拉斯变换的情况,下一段定义的数字梯度的各向同性性质只是对有限数量的旋转增量而言的,这主要取决于所用的近似微分处理的掩模。正如证明的那样,最流行的用于近似梯度处理的掩模仅对水平与垂直边缘给出相同的结果,而梯度处理的各向同性只对 90°的倍数才能保持。这个结果与我们是使用式(3.7.13)还是式(3.7.14)没有什么关系,所以使用两个公式中较简单的算法对实际处理并没有什么影响。

与拉普拉斯情况一样,现在对上述公式定义数字近似方法,并由此得出合适的滤波掩模。为了便于讨论,我们使用图 3.44(a)中的符号来表示 3×3 区域的图像点。例如,若中心点 z_5 表示 $f(x,y)$,那么 z_1 就代表 $f(x-1,y-1)$,以此类推。正如 3.7.1 节中提到的,满足该节规定条件的一阶微分最简单的近似处理就是 $G_x=(z_8-z_5)$和 $G_y=(z_6-z_5)$。先前的数字图像处理中,由 Robert[1965]提出的另两种定义使用了交叉差分算法:

$$G_x=(z_9-z_5) \text{和} G_y=(z_8-z_6) \tag{3.7.15}$$

如果我们选用式(3.7.13)，可以按如下所示计算梯度：

$$\nabla f = [(z_9 - z_5)^2 + (z_8 - z_6)^2]^{\frac{1}{2}} \tag{3.7.16}$$

如果我们使用绝对值，并将式(3.7.15)代入式(3.7.14)，就给出了梯度的近似算法：

$$\nabla f \approx |z_9 - z_5| + |z_8 - z_6| \tag{3.7.17}$$

这个公式可以通过图 3.44(b)与(c)所示的两个掩模得以实现，这些掩模称为 Roberts 交叉梯度算子。

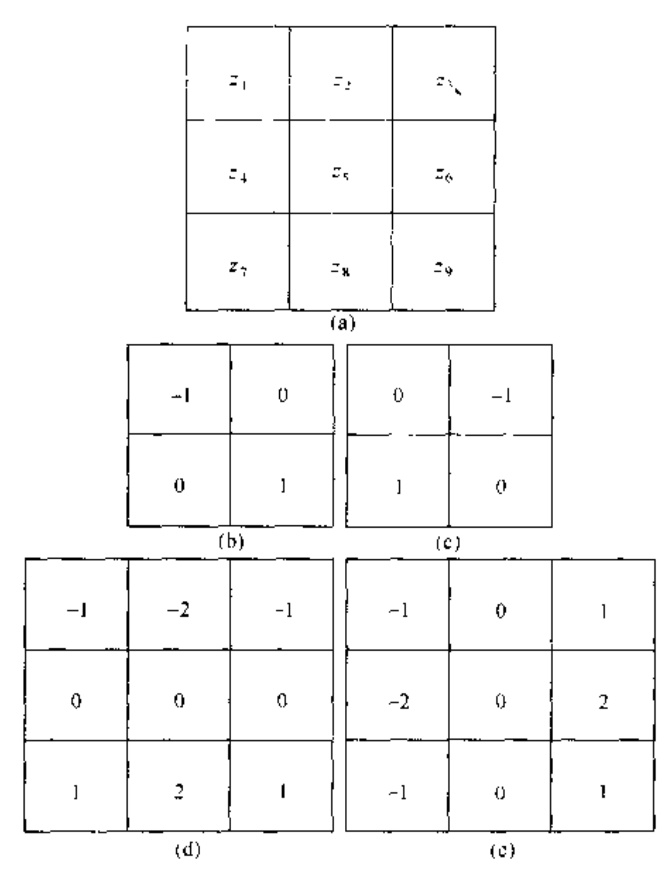


图 3.44　图像的 3×3 区域(z 是灰度值)和用来计算标记为 z_5 的点梯度的掩模。正如所期望的微分算子那样，所有掩模的系数之和为0

偶数尺寸的掩模并不好用。我们感兴趣的是尺寸为 3×3 的最小滤波器掩模。还是在点 z_5，使用绝对值并使用 3×3 掩模的近似结果为：

$$\nabla f \approx |(z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3)| + |(z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7)| \tag{3.7.18}$$

在 3×3 图像区域中，第三行与第一行间的差接近于 x 方向上的微分，同样，第三列与第一列间的差接近于 y 方向上的微分。图 3.44(d)和(e)所示的处理掩模称为 Sobel 算子，它可通过式(3.5.1)的机理实现式(3.7.18)。使用权重 2 的思想是，通过突出中心点的作用而达到平滑的目的(这一点将在第 10 章中做详细介绍)。我们注意到，图 3.44 中所示的所有处理掩模中的系数总和为 0，这表明灰度恒定区域的响应为 0，正如微分算子的期望值那样。

例 3.14　用于边缘增强的梯度处理

梯度处理经常用于工业检测、辅助人工检测缺陷，或者是更为通用的自动检测的预处理。

我们将在第10章和第11章中对此做更多的介绍。为了解释梯度处理能用来突出图像中的小缺陷并能去除变化缓慢的背景特点,我们考虑一个例子。在这个特殊例子中,增强用于自动检测的预处理,而不是视觉分析。

图3.45(a)是一幅隐形眼镜的光学图像,显示了用所设计的光装置突出隐形眼镜的缺陷,例如,在隐形眼镜边缘类似时钟4点和5点处的两个边缘缺陷。图3.45(b)是原始图像用式(3.7.14)与图3.44(d)与(e)中的两个Sobel掩模处理得到的梯度图像。在该图像中,边缘缺陷清晰可见,但是,又增加了一个优点,即灰度不变或变化缓慢的底纹部分被去除了,这样使得缺陷更为突出,且大大简化了自动检测的计算任务。还注意到,梯度处理突出了小斑点,而它们在灰度图像中是看不见的(像这样的小斑点可以是外来物、溶解过程中的气泡、隐形眼镜中的小缺陷)。在灰度平坦区域中增强小突变的能力是梯度处理的另外一项重要特性。

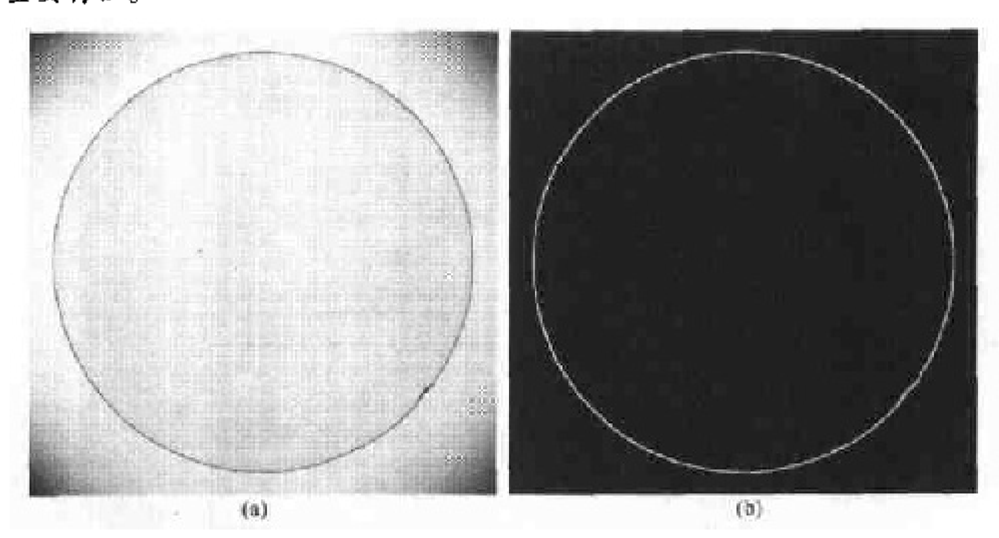


图3.45　(a)隐形眼镜的光学图像(注意,在边缘处类似时钟4点和5点处的缺陷),(b)Sobel梯度(原图像由Perceptics公司的Pete Sites先生提供)

3.8　混合空间增强法

除了一些特殊情况,用3.6.1节中提到的阈值法对混合模糊图像进行处理,到此为止,我们主要关注的还是单一的图像增强法。习惯上,为了实现一个满意的结果,对给定的图像增强目标需要应用多种互补的图像增强技术。在本节,我们通过一个例子来说明怎样将本章中提到的多种图像增强方法结合起来以完成复杂的增强任务。

图3.46(a)所示的图像是一幅人体骨骼核扫描图,它常被用来检查人体疾病,如感染和肿瘤。我们的目标是通过图像锐化突出骨骼的更多细节来增强图像。由于图像灰度的动态范围很窄并且伴随着很高的噪声,所以很难对其进行增强。对此,我们采取的策略是:首先用拉普拉斯变换突出图像中的小细节,然后用梯度法突出其边缘。平滑过的梯度图像将用于掩蔽拉普拉斯图像(见3.4节有关掩蔽的讨论),对此,稍后会做一简短说明。最后,我们通过灰度变换来扩展图像的灰度动态范围。

图3.46(b)显示了用图3.39(d)的掩模得到的原始图像的拉普拉斯变换结果。该图像用与图3.40相同的技术标定过(只为显示)。然后,简单地将图3.46(a)与(b)相加就可以得到一

幅经过锐化处理的图像，它是通过式(3.7.5)的第二行实现的(我们在这里用到中心系数为正的掩模)。只要看一下图 3.46(b)中的噪声水平，就可以预料到，如果将图(a)和(b)相加会导致相当多的噪声，这一点可以在图 3.46(c)中得到证实。于是立刻想到了，减少噪声的方法是使用中值滤波器。然而，中值滤波器是一种非线性滤波器，它有可能改变图像的性质。这在医学图像处理中是不能接受的。

另一种方法就是使用原始图像梯度变换的平滑版本形成的一个掩模。这种方法的动机很直接，并且是基于 3.7.1 节中讲过的一阶和二阶微分性质。拉普拉斯变换作为一种二阶微分算子在图像细节的增强处理方面有明显的优点。但拉普拉斯变换与梯度变换相比会产生更多的噪声。其中那些位于平滑区内的噪声非常显眼而令人讨厌。梯度变换在灰度变化的区域(灰度斜坡或阶梯)的响应要比拉普拉斯变换更为强烈，而梯度变换对噪声和小细节的响应要比拉普拉斯变换弱，而且可以通过均值滤波器对其进行平滑处理进一步降低。这时，对梯度图像进行平滑处理并用拉普拉斯图像与它相乘。在这种情况下，可以将平滑化的梯度变换看成一个模板图像。处理后的结果在灰度变化强的区域仍然保留细节，而在灰度变化相对平坦的区域减少噪声。这种处理可以粗糙地看成将拉普拉斯变换与梯度变换的优点相结合。结果被加到原始图像上就可以得到最终的锐化图像，甚至可用于提升滤波。

图 3.46(d)显示了原始图像经过 Sobel 梯度变换的结果，使用了式(3.7.14)。公式中的分量 G_x 和 G_y 分别通过图 3.44(d)和(e)中的掩模获得。正如我们在 3.7.1 节中讨论的那样，Sobel 图像的边缘要比拉普拉斯图像的边缘突出许多。图 3.46(e)所示的平滑后的梯度图像是用一个 5×5 的均值滤波器获得的。为了以同一方式显示这两幅拉普拉斯图像，两幅梯度图像均被标定过了。由于梯度图像的最小可能值为 0，所以标定的梯度图像的背景呈黑色，而不像标定的拉普拉斯图像那样呈灰色。图 3.46(d)和(e)比图 3.46(b)亮许多的事实再一次证明，带有重要边缘内容的图像梯度值通常要比拉普拉斯图像的高很多。

拉普拉斯与平滑后的梯度图像乘积的结果示于图 3.46(f)。注意，强边缘的优势和可见噪声的减少，这是用一个平滑后的梯度图像掩蔽拉普拉斯变换图像的关键目标。将乘积图像与原始图像相加就得到了图 3.46(g)所示的锐化图像。与原始图像相比，在图像的大部分中，锐化处理过的图像细节的增加都很明显，包括肋骨、脊椎骨、骨盆以及颅骨。单独使用拉普拉斯变换或者梯度变换是不可能达到这种类型的改进效果的。

以上讨论的锐化过程没有影响图像的灰度动态变化范围。我们进行增强处理的最后一步就是扩大锐化图像的灰度动态范围。正如在 3.2 和 3.3 节中详细讨论的，有很多可以完成这一目标的灰度变换函数。从 3.3.2 节处理的结果可以知道，对像本例中这样灰度分布比较暗的图像进行直方图均衡化的效果似乎并不好。直方图归定化可能是一种解决方法，但是，待处理图像的暗灰特性用平方律变换更好。由于我们希望能够扩展灰度范围，所以式(3.2.3)中的 γ 值必须小于 1。通过用这一公式进行一些实验后，得到的结果示于图 3.46(h)，此时 $\gamma=0.5$，$c=1$。将这幅图像与图 3.46(g)相比较，我们看到图 3.46(h)中出现了许多重要的新细节，特别在手腕、手掌、脚踝、脚掌区域。人体的整个骨架也清晰可见，包括手臂骨和腿骨。我们还注意到人体轮廓及人体组织有较弱的清晰度，这是由于通过扩大灰度动态范围显现细节的同时也增大了噪声，但图 3.46(h)与原始图像相比还显示出显著的视觉改进效果。

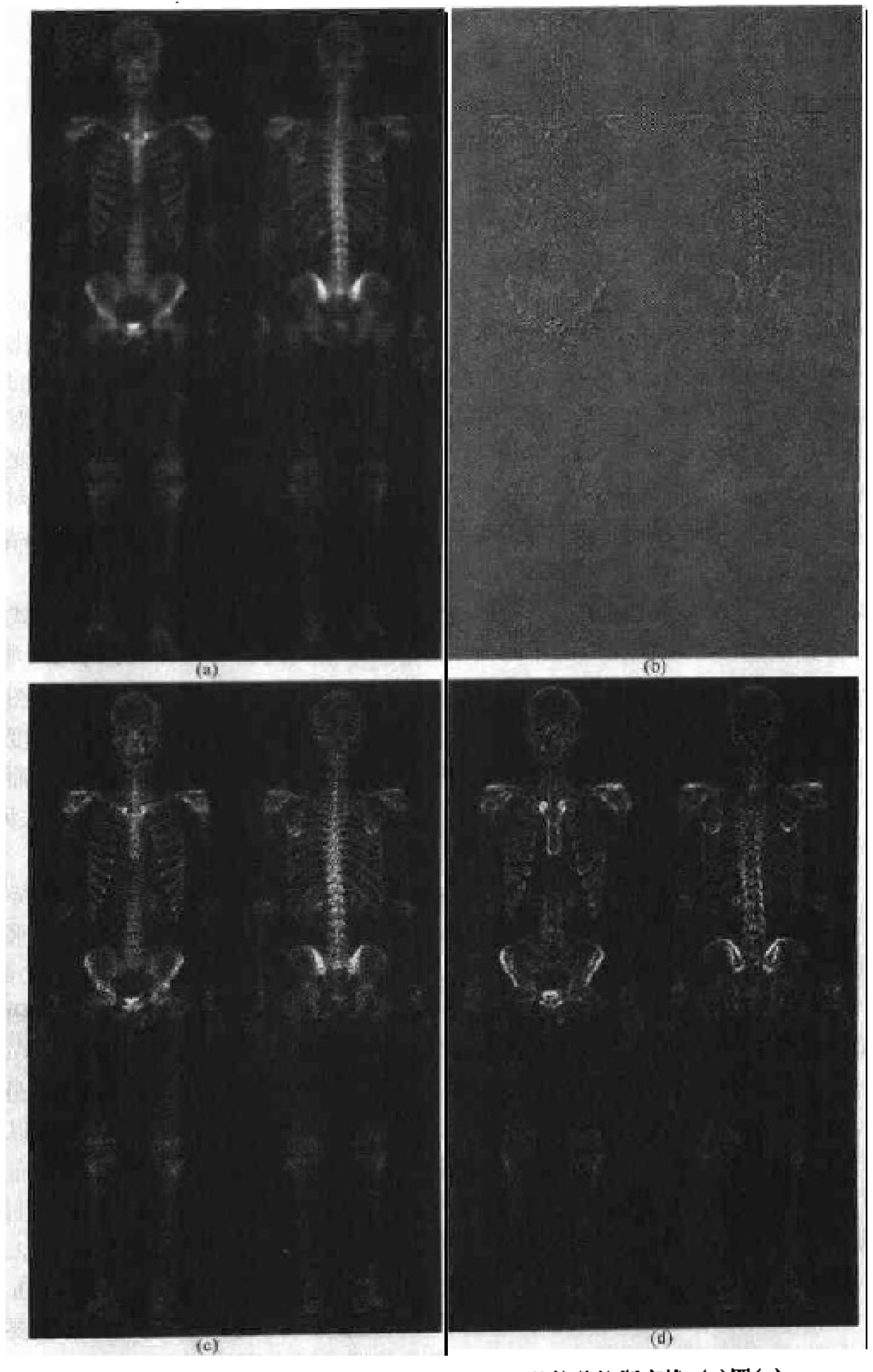


图 3.46 (a)全身骨骼扫描图像,(b)图(a)的拉普拉斯变换,(c)图(a)和图(b)相加得到的锐化图像,(d)图(a)的Sobel处理

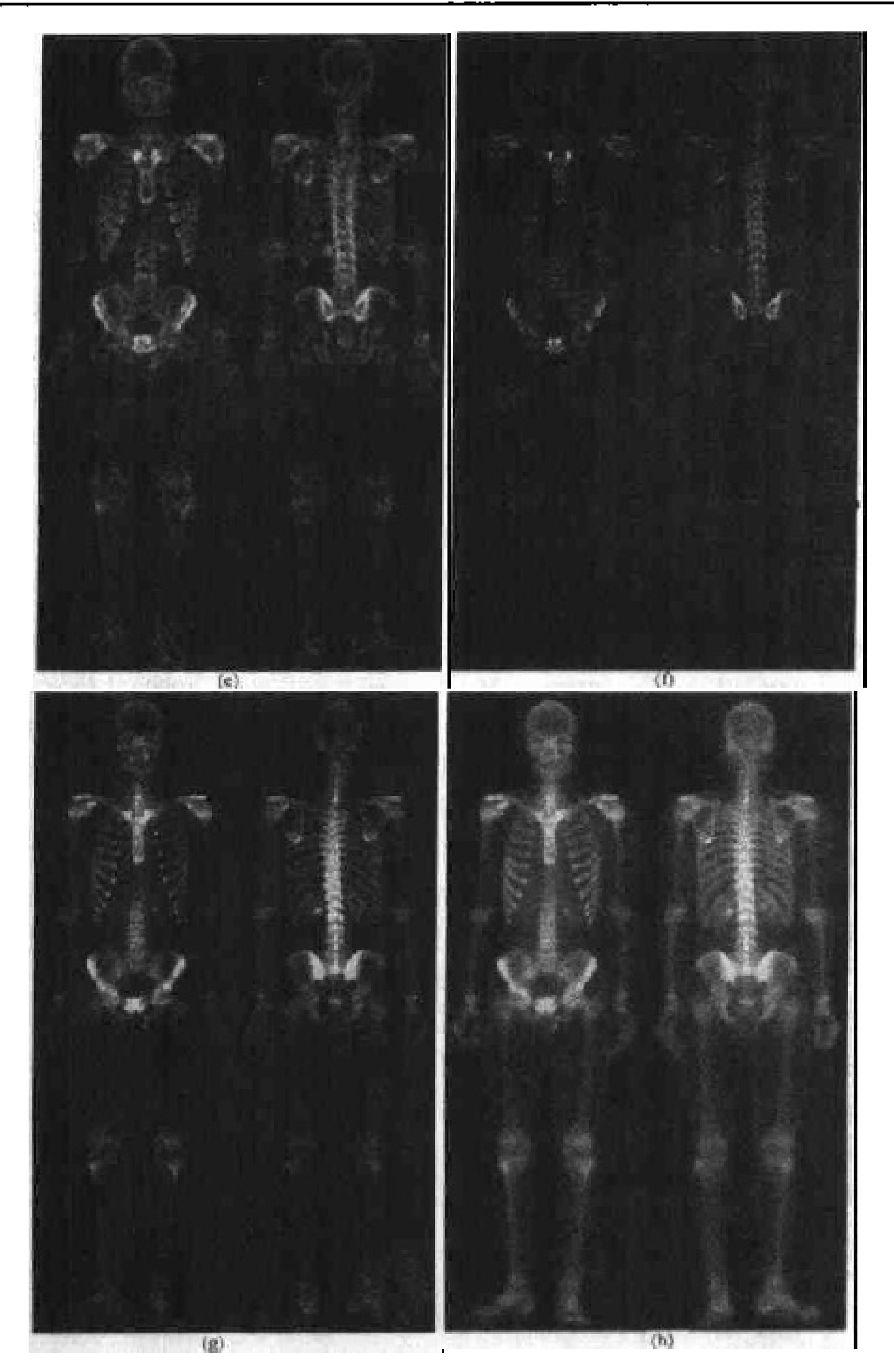


图 3.46(续)　(e)用 5×5 均值滤波平滑的 Sobel 图像，(f)由(c)和(e)相乘形成的掩蔽图像，(g)由(a)和(f)求和得到的锐化图像，(h)对(g)应用幂律变换得到的最后结果。(h)和(g)与(a)比较(原图像由 GE 医学系统提供)

刚才讨论的方法是有代表性的处理形式，为了达到那些只用一种方法所不能实现的效果，可以把多种方法结合起来。如果一些细节作为进一步分析一幅或一系列图像的线索，则在加强它们时，增强图像是非常有用的。而在其他领域，增强处理后的结果可能就是最终的"产品"。在印刷工业、基于图像的产品检测中、法证领域、显微处理领域、监视以及其他领域会发

现,许多情况中增强的主要目标是得到一幅较高视觉细节内容的图像。

小结

本章提供的材料是实践中空间域图像增强有代表性的常用技术。图像处理这一领域是生机勃勃的领域,其新技术和新应用经常可以在一些专业著作和新产品发布中看到。因此,本章中包含的题目都是一些基础材料,可以作为理解增强技术发展状况以及在该领域进一步学习的基础。除了图像增强之外,本章还介绍了一些概念,比如用空间掩模滤波的概念,它在本书的其他章节中还会用到。在接下来的一章中,将在频率域综合讨论图像增强。通过这两章的学习,读者可以在图像处理中所使用的术语和最基础的工具方面打下一个坚实的基础。在图像增强中介绍的这些工具将有助于理解如何处理数字图像。

参考资料

3.1 节中的材料来自 Gonzalez [1986]一书。3.2 节中材料的辅助读物可以参见 Schowengerdt [1983], Poyton [1996]以及 Russ [1999]。还可以参阅 Tsujii 等人 [1998]的论文,这被认为是关于图像显示的最佳材料。直方图处理方面的早期参考书有 Hummel [1974], Gonzalez 和 Fittes [1977]以及 Woods 和 Gonzalez [1981]。Stark [2000] 对自适应对比度增强的直方图均衡处理给出了重要概括。对比度增强的其他方法由 Centeno 和 Haertel [1997] 以及 Cheng 和 Xu [2000]进行了示例说明。对于理想图像模型的增强处理可见 Highnam 和 Brady [1997]的著作。作为局部直方图均衡化处理方法的补充可以参考 Caselles 等人 [1999]以及 Zhu 等人 [1999]的材料。对于图像增强的局部统计实现和应用可以参考 Narendra 和 Fitch [1981]的著述。Kim 等人 [1997] 介绍了梯度处理与局部统计处理相结合的图像增强技术。

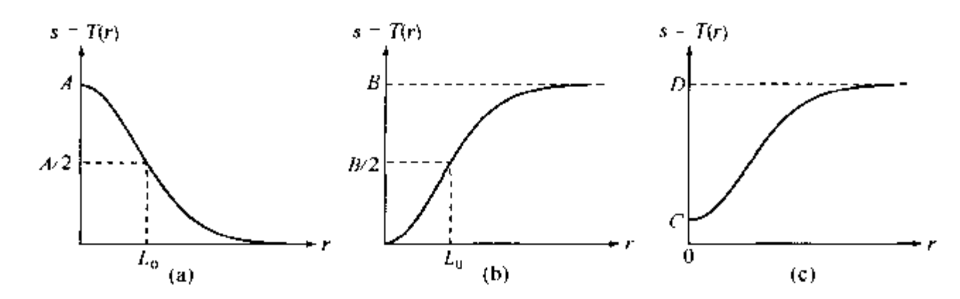
图像的减法处理(3.4.1 节)是一种广泛使用的用于变化检测的普通图像处理工具。正如在那一节中所提到的,数字图像相减处理的主要应用之一是掩模模式射线照相术。这种处理中病人移动是最大的问题,因为移动会使图像变得模糊不清。在图像相减时移动的问题近年来受到了很大的重视,正如 Meijering 等人 [1999]的综述文章中解释的那样。利用图像求平均减少噪声的方法(见 3.4.2 节)由 Kohler 和 Howell [1963]首次提出。可以参考 Peebles [1993]关于随机变量之和的方差与均值的期望值部分。

此外,对于线性空间滤波器及其实现,可以参考 Umbaugh [1998], Jain [1989],以及 Rosenfeld 和 Kak [1982]的材料。Rank-order 滤波器在这些参考材料中也有讨论。Wilburn [1998]也概括地介绍了 Rank-order 滤波器。在 Pitas 和 Venetsanopoulos [1990]的著作中谈到了中值滤波器和其他非线性滤波器。在《IEEE 图像处理学报》[1996]的专刊中介绍了非线性图像处理的内容。有关高提升滤波的内容可以在 Schowengerdt [1983]中找到。在讨论图像复原(第 5 章)和边缘检测(第 10 章)时,我们还会遇到本章中介绍的许多空间滤波器。

习题

3.1 α 为正常数的指数式 $e^{-\alpha r^2}$ 对于构造灰度平滑变换函数是非常有用的。由这个基本函

数开始，构造具有下图形状的变换函数。所示的常数是输入参数，并且提出的变换必须包含这些参数特定的形式（为了答案简单第三条曲线中的 L_0 不是所要求的参数）。



3.2 ★(**a**) 试求出实现示于图 3.2(a)的对比度展宽变换的连续函数。此函数不仅包含参数 m，而且还包括参数 E，以便于控制灰度值由低向高转化时的函数斜率。该函数应归一化，以使它的最小值和最大值分别为 0 和 1。

(**b**) 作为参数 E 的函数，设计一组变换，m 值固定为 $L/2$，L 是图像中灰度的级数。

(**c**) 为使函数如图 3.2(b)的函数那样有效地执行，s 的最小值是什么？换句话说，你的函数与图 3.2(b)可以不同。它仅仅有"产生二值图像"这一相同的处理结果。假定用 8 比特图像进行处理，并使 $m = 128$。另外，令 C 为在你使用的计算机中所能表示的最小正数。

3.3 提出一组能够产生 8 比特单色图像所有单独位平面的灰度分层变换（例如，变换函数 $T(r)$，当 r 在[0,127]范围内时，$T(r) = 0$，而当 r 在[128,255]范围内时，$T(r) = 255$，此函数可以产生一幅 8 比特图像的第 7 位平面图像）。

3.4 ★(**a**)通常，如果将低阶比特面设为零值，对一幅图像的直方图有何影响？

(**b**)如果将高阶比特面设为零值将对直方图有何影响？

★3.5 试解释为什么离散直方图均衡化技术一般不适用于平坦的直方图？

3.6 假定我们对一幅数字图像进行直方图均衡化处理，试说明：第二次直方图均衡化处理的结果与第一次直方图均衡化处理的结果相同。

3.7 在实际应用中，将输入图像的直方图模型化为高斯概率密度函数，其概率密度函数形式为：

$$p_r(r) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(r-m)^2}{2\sigma^2}}$$

其中 m 和 σ 分别是高斯 PDF 的平均值与标准差。具体处理方法是将 m 和 σ 看做给定图像的平均灰度级和对比度，试求出直方图均衡化的变换函数。

★3.8 假设图像的像素值是连续的，用一个例子说明，有可能存在这样的情况，即由式(3.3.4)给出的变换函数满足 3.3.1 节中的条件(a)和(b)，但其反变换却不能满足单值条件。

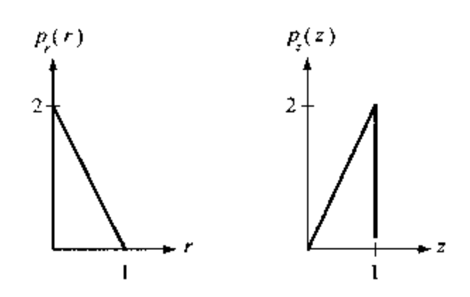
3.9 (**a**) 证明式(3.3.8)中给出的离散变换函数对直方图均衡化处理满足 3.3.1 节中(a)和(b)两个条件。

(**b**) 举例证明式(3.3.9)所表示的离散直方图反变换一般不符合上述两个条件。

★(**c**) 证明在灰度 r_k 不丢失的情况下（$k = 0,1,\cdots,L-1$），式(3.3.9)所表示的离散直方

图反变换才符合3.3.1节中的条件(a)和(b)。

3.10 一幅图像的灰度PDF,$p_r(r)$示于下图。现对此图像进行灰度变换,使其灰度表达式为下面右图的$p_z(z)$。假设灰度值连续,求完成这一要求的变换(r到z)。



★3.11 试提出一种如3.3.3节中讨论的局部增强技术的局部直方图修改方法。

3.12 有两幅图像$f(x,y)$和$g(x,y)$,其直方图分别为h_f和h_g。给出了能确定直方图的条件,并简述在每种情况下,如何得到直方图。

(**a**) $f(x,y)+g(x,y)$

(**b**) $f(x,y)-g(x,y)$

(**c**) $f(x,y)\times g(x,y)$

(**d**) $f(x,y)\div g(x,y)$

3.13 考虑两幅8比特图像,它们的灰度覆盖了整个[0,255]的范围。

(**a**) 讨论反复将图像(b)从图像(a)中减去的最终结果。

(**b**) 如果把图像的顺序颠倒会得到不同的结果吗?

★3.14 图像相减处理常用于工业中,如在生产装配线上检测丢失的元件。这种方法首先要存入一幅"金"图像,即组装正确的图像;然后将相同产品的输入图像减去这幅图像。当然,如果新产品组装正确的话,它们的差为0。如果产品丢失了元件,那么差值图像在该区域与正确图像不同,不是0。你认为在实践中用这种方法应满足什么条件?

3.15 证明式(3.4.4)和式(3.4.5)的正确性。

3.16 在工业应用中,X光摄影法用来检查组合铸件的内部结构。其目的是发现铸件内部是否有缝隙,这些缝隙在图像中一般表现为小气泡。但是,由于铸件材料的性质以及我们使用的X线能量,较高的噪声常会导致检查困难,所以我们决定使用图像均值处理来减小噪声和改进视觉对比度。在计算均值的过程中,为减少时间,在成像中保留固定分块以尽可能保持小的图像数量非常重要。经过数次实践,发现块数为10时噪声的方差明显地减小了。如果成像装置每秒产生30幅图像,那么在成像时保持固定块数而想达到预先要求的铸件噪声方差的减小量,需要多少时间?假设图像中的噪声为非相关的且其均值为0。

3.17 线性空间滤波处理要求在整幅图像中移动掩模的中心点,在每个处理区域中,计算掩模系数与该区域相应像素值乘积的总和(参见3.5节)。在低通滤波器中,所有的系数和为1,我们使用所谓的盒滤波法或移动均值算法(这种方法一次只更新掩模计算的一部分,并从一部分到另一部分顺次进行)。

★(**a**) 对一个$n\times n$的滤波器公式化这样一个算法,说明涉及的计算规律,以及围绕图像移动掩模时所用的扫描序列。

(b) 使用大规模处理的运算量与盒式滤波算法执行的运算量的比值称为"计算优势"。在本例中求出计算优势并作为 n 的函数，其中 $n>1$。由于这两种处理方法的标定系数都是 $1/n^2$，因此，在获得计算优势时，根本不用考虑它的影响。假设图像有为 0 的外部轮廓，这个轮廓足够厚，完全可以忽略它在处理过程中的影响。

3.18 讨论用一个 3×3 低通空间滤波器反复对一幅数字图像处理的结果，可以不考虑边界的影响。

3.19 ★(a) 在 3.6.2 节中谈到，分布在图像背景上的孤立的亮和暗的像素团块，当它们小于中值滤波器区域的一半时，经过中值滤波器处理后会被滤除(被其邻值同化)。假定滤波器尺寸为 $n\times n$，n 为奇数，请解释这种现象的原因？

(b) 考虑一幅有不同像素团块的图像，假设在一个团块内的所有点都比背景亮或者暗(但不是同时既比背景亮又比背景暗)，并且每个团块的尺寸不大于 $n^2/2$。试求当 n 符合什么条件时，有一个或多个这样的团块像(a)中所说的那样被分离出来？

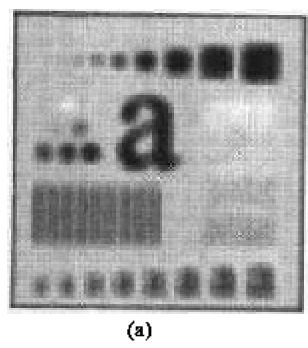
★3.20 (a) 试提出一种过程来求一个 $n\times n$ 邻域的中值？

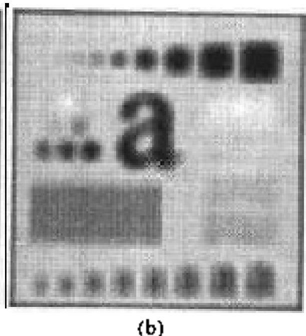
(b) 试提出一种技术，逐像素地移动邻域的中心来更新中值。

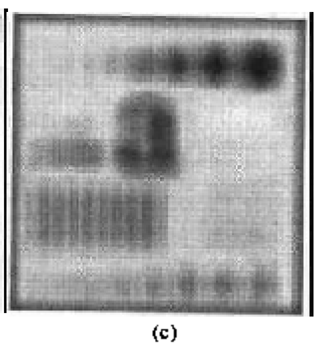
3.21 (a) 在识别的应用领域，文本页通过图 3.2(b)所示的阈值变换函数简化为二值图像。这遵循如下过程，即细化字符直到它们成为全"0"背景上的一串"1"。由于有噪声存在，故二值化和细化处理时，导致在连 1 处有缝隙存在，缝隙有 1 到 3 个像素宽。修复缝隙的一种方法是，对二值图像使用均值掩模来模糊它，这样会在缝隙间桥接非零像素。试求出能执行该任务的均值掩模的最小尺寸？

(b) 桥接缝隙后，为了转换回二值形式，要进行阈值处理。在(a)中得出的答案中，完成这一任务且不产生断线所要求的最小阈值是什么？

★3.22 以下的三幅图像是分别通过 $n=23, 25$ 和 45 的方形均值掩模处理后的模糊图像。图(a)和(c)中左下角的垂直竖条被模糊了，但竖条与竖条之间的分割仍然很清楚。但图(b)中的竖条却已经融入了整幅图像，尽管产生这幅图像的掩模要比处理图像(c)的小得多，请解释这一现象。


(a)


(b)


(c)

3.23 考虑如图 3.36 所示的应用，即消除图像中比 $q\times q$ 像素大小的方形包围的物体小的目标。假设想要将目标的平均灰度减少为原来平均灰度的 1/10。用这种方法，那些目标可以接近背景灰度并用门限法消除。给出平均掩模的最小尺寸(奇数)，该掩模仅对整幅图像处理一次就可将平均灰度级减少到所希望的程度。

3.24 在给定的应用中,一个均值掩模被用于输入图像以减少噪声,然后再用一个拉普拉斯掩模来增强图像中的小细节,如果将这两个步骤交换一下,结果是否会相同?

★**3.25** 证明如式(3.7.1)所示的拉普拉斯变换是各向同性的(旋转不变)。需要下列轴旋转 θ 角的坐标方程:

$$x = x'\cos\theta - y'\sin\theta$$
$$y = x'\sin\theta + y'\cos\theta$$

其中(x,y)为非旋转坐标,而(x',y')为旋转坐标。

3.26 试求一个 3×3 的反锐化掩模来对一幅图像进行一次掩模处理。

★**3.27** 使用式(3.7.4)给出的拉普拉斯变换的定义,证明将一幅图像减去其相应的拉普拉斯图像等同于对图像做反锐化掩模处理。

3.28 (a)证明式(3.7.13)给出的梯度值是一种各向同性的操作过程(见习题3.25)。
(b)证明如果梯度用式(3.7.14)进行计算将使其失去各向同性的性质。

3.29 CCD电视摄像机用于每天24小时、每月30天的对同一区域进行的长期观测研究。每5分钟拍取一次数字图像并传送到中心位置。场景的照明,白天为自然光,晚上为人造光,没有无照明的时间,因此,总可以获得一幅图像。由于照明的范围总是在摄像机的线性工作区,因此,摄像机本身并不需要使用任何补偿装置。另外,决定使用数字技术对图像进行后处理并进行归一化,这样,图像与恒定照明是等效的。对此,设计一种方法。可以在实验室内使用希望的任何方法,但要在设计中明确列出所做的所有假设。

第4章 频率域图像增强

尽管在前一章中已着重讲述了图像增强的空间技术,但如果不了解傅里叶变换和频率域在图像处理中如何应用的基本知识,要彻底地理解这一领域也是不太可能的。你不必是一位信号处理专家就可对这些主题有深刻理解。关键在于要集中精力于原理和与数字图像处理有关的内容上。表示法——通常是初学者感到困难的根源——在此章中重点阐述,并强调图像特征与表示这些特征的数学工具之间的联系。本章主要是帮助读者建立对傅里叶变换和频率域的基本理解,以及明确如何将它们应用于图像增强上。而后,在第5章、第8章、第10章和第11章中,将讨论傅里叶变换的其他应用。

本章从傅里叶变换的起源和其应用于数学、科学及工程的多个分支的概况开始简单谈起。然后,介绍傅里叶变换及频率域,阐明表示法以及这些工具对于图像增强十分有用的原因。接下来是与在第3章中讨论的空间平滑和锐化滤波技术(除了在频率域中进行的所有滤波)相对应的部分。还要讨论傅里叶变换对于图像增强的其他应用,最后,用在图像处理中实现傅里叶变换问题的探讨总结本章内容。

4.1 背景

法国数学家傅里叶生于1768年,他被世人铭记的最大贡献记载在1807年的传记中和后来出版于1822年的"La Théorie Analitique de la chaleur"(热分析理论)一书中。此书由Freeman(参见Freeman[1878])在55年后翻译为英文。傅里叶在这个特殊领域的贡献是他指出任何周期函数都可以表示为不同频率的正弦和/或余弦和的形式,每个正弦和/或余弦乘以不同的系数(现在称这个和为傅里叶级数)。无论函数有多么复杂,只要它是周期的,并且满足某些软的数学条件,都可以用这样的和来表示。我们现在认为理所应当,但在它第一次出现的时候,这个革命性的概念被全世界的数学家"纠正"了一个世纪。在那时,在数学思想中函数的规律性是占主导的。基于这样的传统思想,复杂函数可以由简单的正弦和余弦之和来表示的概念根本不直观(如图4.1所示),所以傅里叶的想法遭到怀疑是不足为奇的。

甚至非周期的函数(但是这些领域是在曲线是有限的情况下)也可以用正弦和/或余弦乘以加权函数的积分来表示。在这种情况下的公式就是傅里叶变换,它的应用在大多数实际应用中比傅里叶级数更广泛。用傅里叶级数或变换表示的函数特征可以完全通过傅里叶反过程来重建,不丢失任何信息。这是这些表示法的最重要特征之一,因为它可以使我们工作于"频率域",而且在转换回函数的原始域时不丢失任何信息。

总之,傅里叶级数和变换是解决实际问题的工具,它被广泛地使用并作为基础工具学习。傅里叶最初想法的应用是在热扩散领域,人们考虑用微分方程的公式表示热流动,用这种方法第一次获得了结论。在过去的一个世纪里,尤其是后50年,傅里叶的思想使整个工业和学术界都空前繁荣。在20世纪50年代后期,数字计算的出现和快速傅里叶变换算法的"发明"在信号处理领域产生了巨大变革。这两个核心技术第一次允许对人类本身的特殊信

号和工业的重要信号（从医学监视器和扫描仪到现代电子通信），进行实际处理和有意义的解释。

这里仅处理有限域内的函数(图像),所以傅里叶变换是我们感兴趣的工具。在以下各节的内容中将介绍傅里叶变换和频率域。它显示出傅里叶技术提供了一个有意义的和实际的研究及实现图像增强的主要途径。在某些情况下,这些途径与我们在第3章中阐述的方式相近。而在其他方面,它们也可以作为补充。

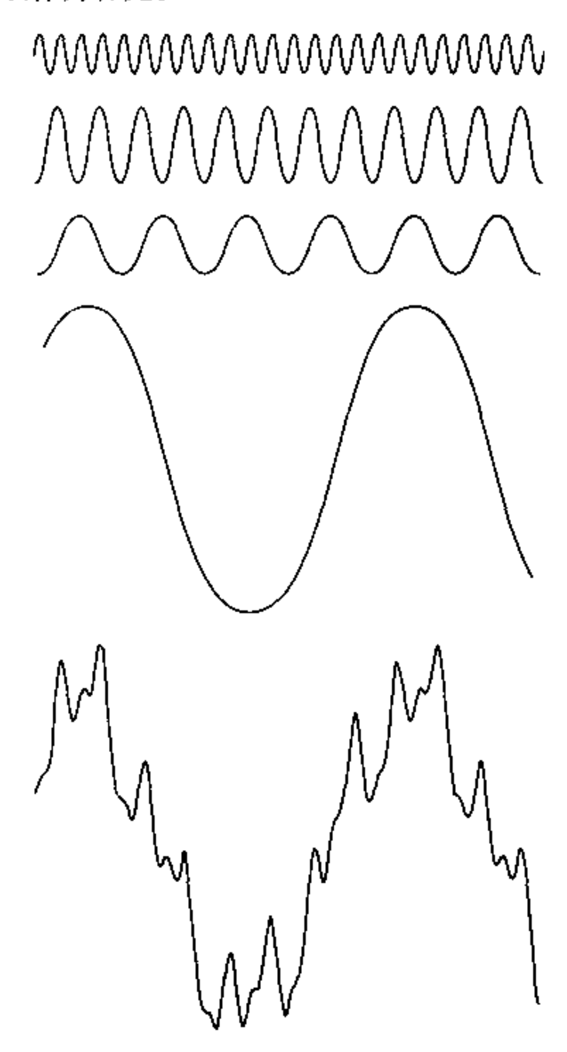


图4.1　在最下面的函数是上面四个函数的和。在1807年,傅里叶的思想——周期函数可以表示为加权的正弦与余弦和的形式——遭到了怀疑

4.2　傅里叶变换和频率域的介绍

本节介绍一维和二维傅里叶变换。重点在连续变换的离散化公式和其一些性质上。

4.2.1　一维傅里叶变换及其反变换

单变量连续函数 $f(x)$的傅里叶变换 $F(u)$定义为等式：

$$F(u) = \int_{-\infty}^{\infty} f(x) e^{-j2\pi ux} dx \tag{4.2.1}$$

其中 $j = \sqrt{-1}$。相反，给定 $F(u)$，通过傅里叶反变换可以获得 $f(x)$：

$$f(x) = \int_{-\infty}^{\infty} F(u) e^{j2\pi ux} du \tag{4.2.2}$$

这两个等式组成了傅里叶变换对。它们指出在前一节中提到的重要事实，即一个函数可以从它的反变换中重新获得。这些等式很容易扩展到两个变量 u 和 v：

$$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y) e^{-j2\pi(ux+vy)} dx dy \tag{4.2.3}$$

类似地，反变换为：

$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v) e^{j2\pi(ux+vy)} du\ dv \tag{4.2.4}$$

我们的兴趣在于离散函数，所以将不停留在这些等式中。然而，在某些情况下，读者会发现利用这些等式比利用它们的离散形式更容易证明二维傅里叶变换的性质。

单变量离散函数 $f(x)$（其中 $x = 0,1,2,\cdots,M-1$）的傅里叶变换由以下等式给出：

$$F(u) = \frac{1}{M}\sum_{x=0}^{M-1} f(x) e^{-j2\pi ux/M} \qquad u = 0,1,2,\cdots,M-1 \tag{4.2.5}$$

离散傅里叶变换（DFT）是这一章中大部分工作的基础。同样，给出 $F(u)$，能用反 DFT 来获得原函数：

$$f(x) = \sum_{u=0}^{M-1} F(u) e^{j2\pi ux/M} \qquad x = 0,1,2,\cdots,M-1 \tag{4.2.6}$$

在傅里叶变换前的 $1/M$ 乘数有时被放置在反变换前。其他时候（非经常），两个等式都乘以 $1/\sqrt{M}$。乘数的位置并不重要。如果使用两个乘数，仅要求必须使乘积结果为 $1/M$。相比它们的重要性，这些等式的确非常简单。

为了计算在式（4.2.5）中的 $F(u)$，首先在指数项中代入 $u=0$，然后，将所有 x 值相加。之后，在指数项中代入 $u=1$，重复对所有 x 的相加。对所有 M 个 u 值重复这一过程，从而可获得完整的傅里叶变换。这个过程花费了将近 M^2 个加法和乘法来计算离散傅里叶变换（减少这个运算数量是一个很重要的课题，将在 4.6 节中讨论）。像 $f(x)$ 一样，变换在数量上也是离散的，有着与 $f(x)$ 相同数量的分量。相似的注释同样适用于反傅里叶变换的计算。

离散变换对的一个很重要的特性是，不像连续的情形，不必关心 DFT 或它的反变换是否存在。离散傅里叶变换和它的反变换总是存在的。这一点可以从将式（4.2.5）或式（4.2.6）相互代入到对方式子中以及利用指数的正交特性（习题 4.1）来得出。可以获得一个指明这两个函数存在的恒等式。当然，当 $f(x)$ 为无限值时，总会有些问题发生，但我们在本书中仅处理有限数值。这些注释都直接适用于二维（和更高维数）函数。因此，对于数字图像处理，离散变换或其反变换的存在不是问题。

频率域的概念在本章和第 3 章被多次提到，从欧拉公式中得到：

$$e^{j\theta} = \cos\theta + j\sin\theta \tag{4.2.7}$$

将此表达式代入式（4.2.5）中，使用公式 $\cos(-\theta) = \cos\theta$，得出：

$$F(u) = \frac{1}{M}\sum_{x=0}^{M-1} f(x)[\cos 2\pi ux/M - j\sin 2\pi ux/M] \tag{4.2.8}$$

其中,$u=0,1,2,\cdots,M-1$。因此,我们看到傅里叶变换的每一项[即对于每个 u 值,$F(u)$的值]由 $f(x)$函数所有值的和组成。$f(x)$的值则与各种频率的正弦值和余弦值相乘。$F(u)$值的范围覆盖的域(u 的值)称为频率域,因为 u 决定了变换的频率成分(x 也作用于频率,但它们相加,对每个 u 值有相同的贡献)。$F(u)$的 M 项中的每一个被称为变换的频率分量。使用术语"频率域"和"频率成分"与"时间域"和"时间成分"没有差别,如果 x 是一个时间变量,可以用它来表示 $f(x)$的域和值。

一个恰当的比喻是将傅里叶变换比做一个玻璃棱镜。棱镜是可以将光分成不同颜色成分的物理仪器,每个成分的颜色由波长(或频率)决定。傅里叶变换可看做"数学的棱镜",将函数基于频率分成不同的成分。当我们考虑光时,讨论它的光谱或频率谱线。同样,傅里叶变换使我们能够通过频率成分来分析一个函数。这是属于线性滤波核心的重要概念。

总之,从式(4.2.5)或式(4.2.8)中可以看出傅里叶变换的成分很复杂。正如在复数的分析中,我们发现有时在极坐标下表示 $F(u)$很方便:

$$F(u)=|F(u)|e^{-j\phi(u)} \tag{4.2.9}$$

其中

$$|F(u)|=\left[R^2(u)+I^2(u)\right]^{1/2} \tag{4.2.10}$$

上式称为傅里叶变换的幅度或频率谱,同时

$$\phi(u)=\arctan\left[\frac{I(u)}{R(u)}\right] \tag{4.2.11}$$

上式称为变换的相角或相位谱。在式(4.1.10)和式(4.2.11)中,$R(u)$和 $I(u)$分别是 $F(u)$的实部和虚部。在研究图像增强时,我们主要关心频率谱的性质。在本章后面使用的另一个量是功率谱,它被定义为傅里叶变换的平方:

$$\begin{aligned} P(u) &= |F(u)|^2 \\ &= R^2(u)+I^2(u) \end{aligned} \tag{4.2.12}$$

术语"谱密度"也用来指功率谱。

例 4.1 两个简单一维函数的傅里叶谱

在继续下面的内容之前,先考虑一个简单的一维 DFT 例子会有些帮助。图 4.2(a)显示了一个函数,图 4.2(b)显示了它的傅里叶频率谱。$f(x)$和 $F(u)$都是离散量,图中的点相连是为了便于观看。在此例子中,$M=1\,024$,$A=1$,K 仅为 8 个点。同时注意频率谱的中心在 $u=0$ 处。正如接下来的一节所显示的那样,在进行变换前 $f(x)$被乘以$(-1)^x$ 以完成此图。下面的两个图描述的与前边基本相同,但 K 为 16 个点。值得注意的重要特征为:(1)当曲线下的面积在 x 域加倍时,频率谱的高度也加倍;(2)当函数的长度加倍时,相同间隔下频谱中零点的数量也加倍。傅里叶变换对的这个"对偶"特性在频率域下解释图像处理结果时非常有用。

在式(4.2.5)的离散傅里叶变换中,函数 $f(x)$, $x=0,1,2,\cdots,M-1$,表示从连续的公式中取 M 个样点。记住这一点很重要,这些样点不一定必须选取 x 在$[0,M-1]$范围内的整数值。但在这里它们是任意选取的等间隔点。经常用 x_0(任意定位的)表示序列中的第一个点。被抽样函数的第一个值就是 $f(x_0)$。下一个样点采用固定间隔 Δx,从而得到 $f(x_0+\Delta x)$。第 k 个样点为 $f(x_0+k\Delta x)$,最后一个样点为 $f(x_0+[M-1]\Delta x)$。这样,在离散情况下,当写出

$f(k)$时,可以理解为是 $f(x_0+k\Delta x)$的缩写表示。根据这样的表示法,当处理离散变量时,$f(x)$被理解为:

$$f(x) \triangleq f(x_0+x\Delta x) \tag{4.2.13}$$

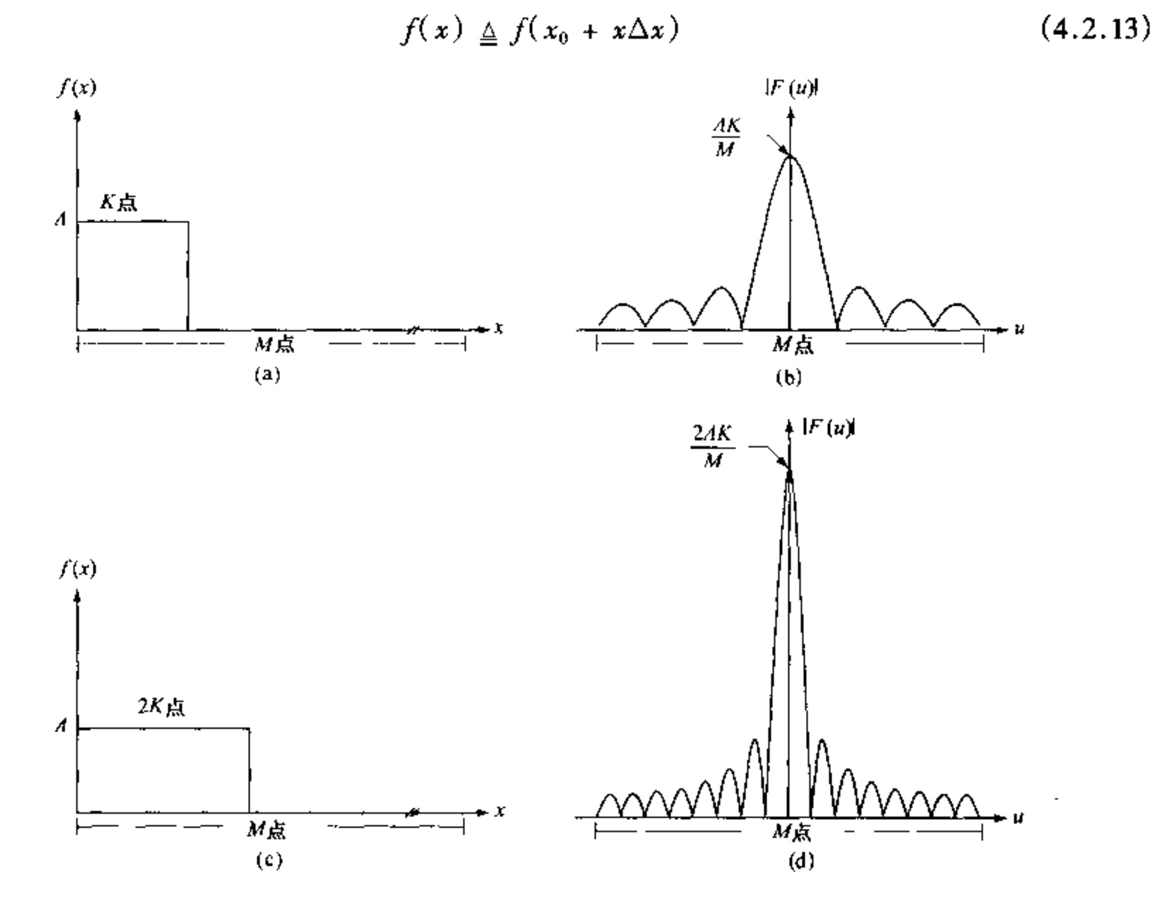


图 4.2　(a)M 点的一个离散函数,(b)它的傅里叶频谱,(c)两倍非零点数目的离散函数,(d)它的傅里叶频谱

变量 u 有相似的解释,但序列通常总是从 0 频率开始。因此,u 值序列为 $0,\Delta u,2\Delta u,\cdots,[M-1]\Delta u$。$F(u)$理解为:

$$F(u) \triangleq F(u\Delta u) \tag{4.2.14}$$

其中,$u=0,1,2,\cdots M-1$。这种缩写大大简化了等式,并容易理解得多。

在图 4.2 中给出了函数和其变换之间的相反关系,显然,Δx 和 Δu 有如下关系:

$$\Delta u = \frac{1}{M\Delta x} \tag{4.2.15}$$

在图像处理中当测量是一个问题时这个关系很有用。例如,在电子显微镜下的应用,图像样品可能分开 1 μm 的间隙,在频率域中某些特性可根据实际样品的结构暗示出来。本书接下来的大部分讨论中,在不考虑特定抽样或其他测量因素的情况下,使用变量 x 和 u。

4.2.2　二维 DFT 及其反变换

一维离散傅里叶变换及其反变换向二维扩展是简单明了的。一个图像尺寸为 $M\times N$ 的函数$f(x,y)$的离散傅里叶变换由以下等式给出:

$$F(u,v)=\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)e^{-j2\pi(ux/M+vy/N)} \tag{4.2.16}$$

像在一维中的情形一样,此表达式必须对 u 值($u=0,1,2,\cdots,M-1$)和 v 值($v=0,1,2,\cdots,N-1$)计算。同样,给出 $F(u,v)$,可以通过反傅里叶变换获得 $f(x,y)$,由表达式给出:

$$f(x,y)=\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}F(u,v)e^{j2\pi(ux/M+vy/N)} \tag{4.2.17}$$

其中,$x=0,1,2,\cdots,M-1$,$y=0,1,2,\cdots,N-1$。式(4.2.16)和式(4.2.17)构成了二维离散傅里叶变换对(DFT)。变量 u 和 v 是变换或频率变量,x 和 y 是空间或图像变量。正如在一维中的情形那样,常量 $1/MN$ 的位置并不重要,有时它在反变换之前。其他时候,它被分为两个相等的常数 $1/\sqrt{MN}$,分别乘在变换和反变换的式子前。

像在前一节中一样定义傅里叶谱、相角和频率谱:

$$|F(u,v)|=\left[R^2(x,y)+I^2(x,y)\right]^{1/2} \tag{4.2.18}$$

$$\phi(u,v)=\arctan\left[\frac{I(u,v)}{R(u,v)}\right] \tag{4.2.19}$$

并且

$$\begin{aligned}P(u,v)&=|F(u,v)|^2\\&=R^2(u,v)+I^2(u,v)\end{aligned} \tag{4.2.20}$$

其中,$R(u,v)$和 $I(u,v)$分别是 $F(u,v)$的实部和虚部。

通常在进行傅里叶变换之前用$(-1)^{x+y}$乘以输入的图像函数。由于指数的性质,很容易看出(参见 4.6 节):

$$\Im[f(x,y)(-1)^{x+y}]=F(u-M/2,v-N/2) \tag{4.2.21}$$

其中 $\Im[\cdot]$表示引文中的傅里叶变换。这个等式说明 $f(x,y)(-1)^{x+y}$傅里叶变换的原点[即 $F(0,0)$]被设置在 $u=M/2$ 和 $v=N/2$ 上。换句话说,用$(-1)^{x+y}$乘以 $f(x,y)$将 $F(u,v)$原点变换到频率坐标下的($M/2$,$N/2$),它是二维 DFT 设置的 $M\times N$ 区域的中心。我们将此频率域的范围指定为频率矩形,它从 $u=0$ 到 $u=M-1$,从 $v=0$ 到 $v=N-1$(u 和 v 是整数)。为了确保移动后的坐标为整数,要求 M 和 N 为偶数。当在计算机中使用傅里叶变换时,总和的范围为 u 从 1 到 M,v 从 1 到 N。实际的变换中心将为 $u=(M/2)+1$ 和 $v=(N/2)+1$。

从式(4.2.16)得到$(u,v)=(0,0)$的变换值为:

$$F(0,0)=\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y) \tag{4.2.22}$$

即 $f(x,y)$的平均值。换句话说,如果 $f(x,y)$是一幅图像,在原点的傅里叶变换即等于图像的平均灰度级。因为在原点处常常为零,$F(0,0)$有时称做频率谱的直流成分。这个术语来自电子工程领域,其中"dc"表示直流(即,零频率的电流)。

如果 $f(x,y)$是实函数,它的傅里叶变换必然为对称的,即:

$$F(u,v)=F^*(-u,-v) \tag{4.2.23}$$

其中"$*$"表示对于复数的标准共轭操作。由此,它遵循:

$$|F(u,v)|=|F(-u,-v)| \tag{4.2.24}$$

其中,傅里叶变换的频率谱为对称的。共轭对称和前面讨论的中心对称的性质简化了频率域

内循环对称滤波器的技术条件,此内容将在下面的部分中讨论。

最后,如在一维中的情况,空间域和频率域抽样点之间的关系如下所示:

$$\Delta u = \frac{1}{M\Delta x} \tag{4.2.25}$$

并且

$$\Delta v = \frac{1}{N\Delta y} \tag{4.2.26}$$

这些变量的意义解释与 4.2.1 节中给出的一维变量相同。

例 4.2　一个简单二维函数的中心谱

图 4.3(a)显示了在 512×512 像素尺寸的黑色背景上叠加一个 20×40 像素尺寸的白色矩形。此图像在进行傅里叶变换的计算之前被乘以$(-1)^{x+y}$,从而可以使频率谱关于中心对称,如图 4.3(b)所示(注意两张图的位置、标志和轴的原点。在图像和相关的傅里叶谱中的讨论都遵循此约定)。在图 4.3(b)中,u 方向谱的零点分隔恰好是 v 方向零点分隔的两倍。这却相反地符合图像中 1:2 的矩形尺寸比例。在显示之前频率谱用式(3.2.2)中的对数变换处理以增强灰度级细节。变换中使用 $c=0.5$ 的值可以降低整体强度。在本章中显示的多数傅里叶频率谱都用对数变换进行了相似的处理。

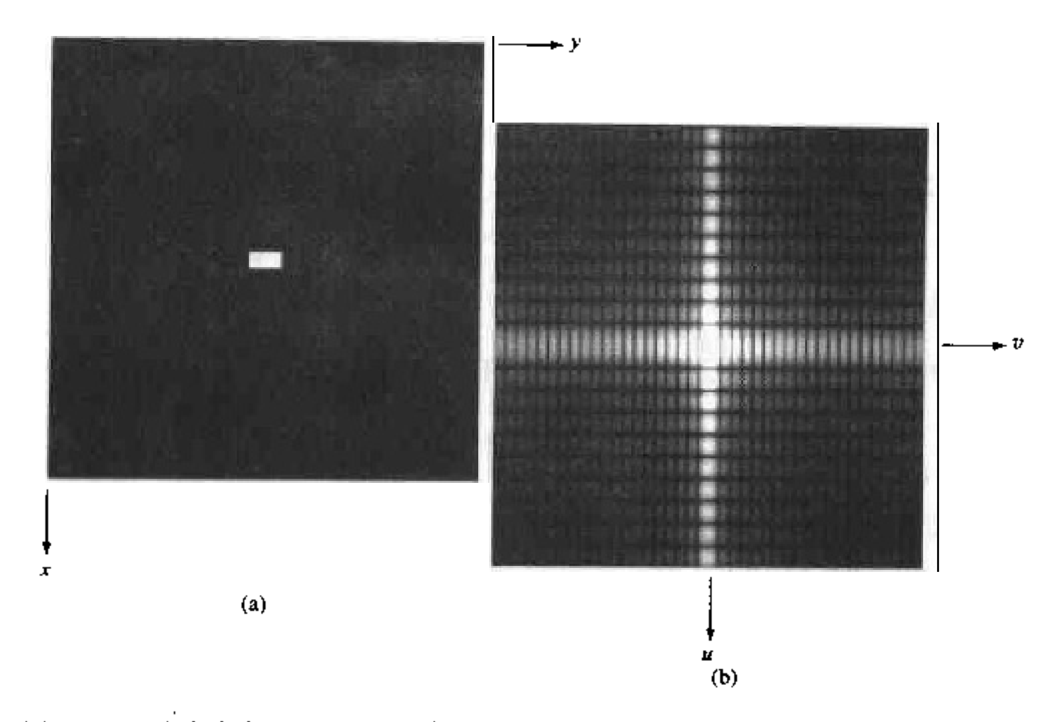


图 4.3　(a)在大小为 512×512 的黑色背景上叠加一个尺寸为 20×40 的白色矩形的图像,(b)与图4.2相比,应用了式(3.2.2)中给出的对数变换后显示的中心傅里叶谱

4.2.3　频率域滤波

在前两节中,频率域仅是由傅里叶变换和频率变量(u,v)定义的空间。此节中,将联系图像处理揭示频率域的深刻含义。

频率域的基本性质

从观察式(4.2.16)开始,每个 $F(u,v)$项包含了被指数项修正的 $f(x,y)$的所有值。因此,除了特殊情况,一般不可能建立图像特定分量和其变换之间的直接联系。然而,一般文献通常会有关于傅里叶变换的频率分量和图像空间特征之间联系的阐述。例如,既然频率与变化率直接相关,直观上要将傅里叶变换的频率与图像中的强度变化模式联系起来并不困难。在前面的章节中显示了变化最慢的频率成分($u=v=0$)对应一幅图像的平均灰度级。当从变换的原点移开时,低频对应着图像的慢变化分量,例如,一幅房间的图像,墙和地板可能对应平滑的灰度分量,当我们进一步移开原点时,较高的频率开始对应图像中变化越来越快的灰度级。这些是物体的边缘和由灰度级的突发改变(如噪声)标志的图像成分。

例 4.3 一幅图像及显示某些重要特征的傅里叶谱

一个例证将有助于确定这些想法。图 4.4(a)中所示图像是一幅集成电路的扫描电子显微镜图像,放大将近2500倍。除了设备本身有趣的结构外,还注意到两个主要的特征:大约成 ±45°的强边缘和两个因热感应不足而产生的白色氧化突起。图4.4(b)所示的傅里叶频谱显示了沿着 ±45°方向对应于刚刚提及的边缘突起的部分。沿着垂直轴仔细观察,可以看到在轴偏左的部分有垂直成分。这是由氧化突起的边缘形成的。注意在偏离轴的角度,频率成分如何对应长的白色元素的水平位移,并且注意在垂直频率成分中的零点如何对应氧化突起的狭窄垂直区域。

这是典型的建立频率和空间域联系的例子。正如在本章所示的,即使是这些总体的联系类型,以及前面提到的频率与图像灰度级变化率之间的联系,也可以产生非常有用的增强结果。

在频率域中的滤波基础

在频率域中的滤波是简单明了的。它包含如下步骤:

1. 用$(-1)^{x+y}$乘以输入图像来进行中心变换,如式(4.2.21)所示。
2. 由(1)计算图像的 DFT,即 $F(u,v)$。
3. 用滤波器函数 $H(u,v)$乘以 $F(u,v)$。
4. 计算(3)中结果的反 DFT。
5. 得到(4)中结果的实部。
6. 用$(-1)^{x+y}$乘以(5)中的结果。

$H(u,v)$被称为滤波器(常用术语为"滤波器传递函数")的原因是它在变换中抑制某些频率但其他频率不受影响。类似于日常生活中的筛子,它可以严格按照尺寸使一些物体通过而同时阻止其他物体。

在等式的形式下,用 $f(x,y)$代表步骤(1)中的输入图像,而用 $F(u,v)$表示它的傅里叶变换。如下给出其输出图像的傅里叶变换:

$$G(u,v) = H(u,v)F(u,v) \tag{4.2.27}$$

H 和 F 的相乘涉及二维函数,并在逐元素的基础上定义。即,H 的第一个元素乘以 F 的第一个元素,H 的第二个元素乘以 F 的第二个元素,以此类推。总之,F 中的元素为复数,但我们

在本书中处理的典型滤波器为实数。在这种情况下，H 的每一个分量乘以 F 中相应部分的实部和虚部。这种滤波器称为"零相移"滤波器。正如名字中提示的，这些滤波器不改变变换的相位，从式(4.2.19)可以看出因为值相等，所以实部和虚部的乘数可以抵消。

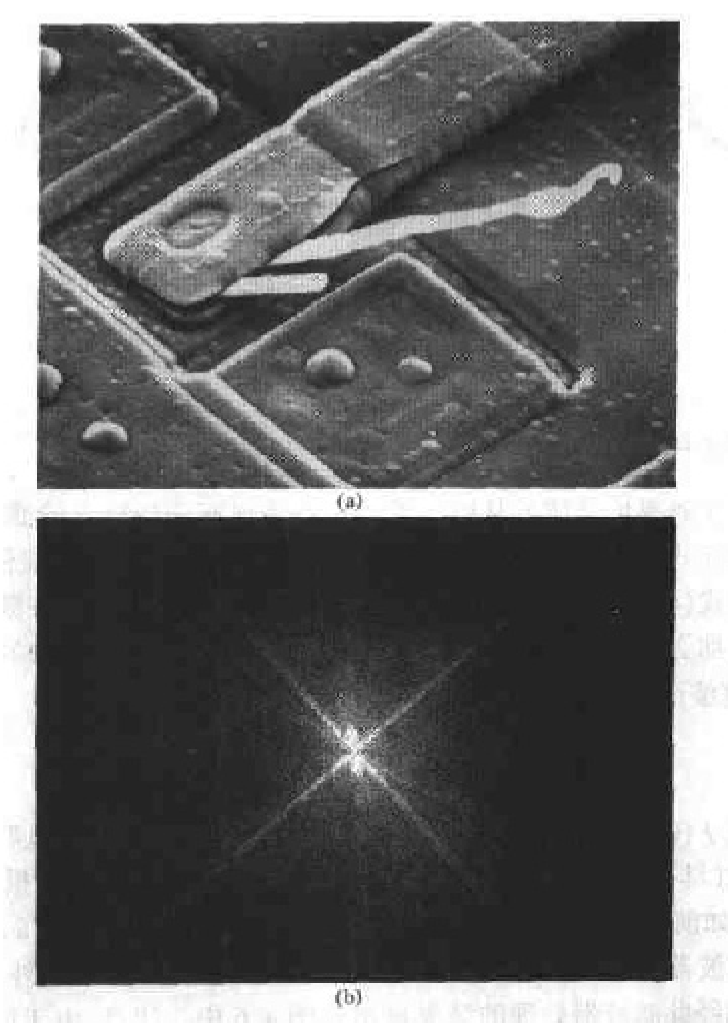


图 4.4　(a)受损的集成电路 SEM 图像，(b)(a)的傅里叶谱
(原图像由加拿大安大略省哈密尔顿市 McMaster 大学材料研究所的 J.M.Hudak 博士提供)

被滤波的图像可以从 $G(u,v)$的反傅里叶变换中得到：

$$\text{被滤波的图像} = \Im^{-1}[G(u,v)] \tag{4.2.28}$$

最终的图像通过取结果的实部并且乘以$(-1)^{x+y}$以取消输入图像的乘数来获得。总体来说，反傅里叶变换较复杂。然而，当输入图像和滤波器函数都为实函数时，反傅里叶变换的虚部即为 0。实际上，由于计算的舍入误差，反 DFT 一般有寄生的虚部成分，这些成分可以忽略。

包括前处理和后处理的更普遍形式的滤波过程框图总结在图 4.5 中。除了$(-1)^{x+y}$过程之外，其他过程的例子可能包括将输入图像向其最接近的偶数维数转换(要求有合适的变换中心)、灰度级标定、输入向浮点的转换、输出向 8 比特整数格式的转换等。多滤波步骤和其他前后处理函数都是有可能的。在这个基本主题上会有许多变化。需要记住的关键点是，滤波过程通过滤波器函数以某种方式来修改图像变换，然后通过取结果的反变换来获得处理后的输出图像。

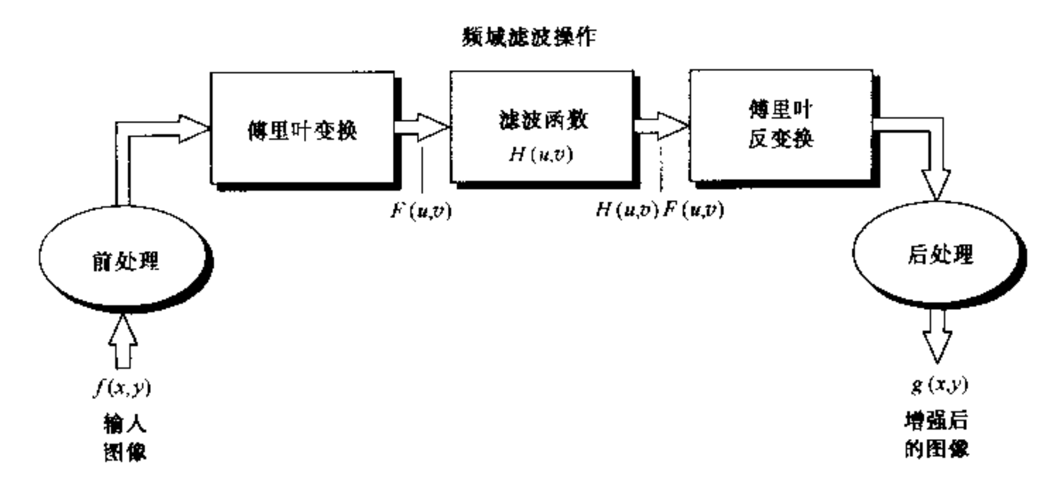


图 4.5 频域滤波的基本步骤

一些基本的滤波器及其性质

目前已经建立了频率域滤波的基础。下一步就是观察一些特殊的滤波器,看它们是如何作用于图像的。前面式(4.2.22)的讨论很好地把我们引入到了滤波的实例。假定希望图像的平均值为零。根据式(4.2.22),图像的平均值由 $F(0,0)$ 给出。如果在频率域中设置此项为零,并进行反变换,那么结果图像的平均值将为零。假定变换如式(4.2.21)中所讨论的那样集中在中心,则可以用滤波函数乘以 $F(u,v)$ 的所有值来完成此操作。

$$H(u,v)=\begin{cases}0 & (u,v)=(M/2,N/2)\\ 1 & \text{其他}\end{cases} \tag{4.2.29}$$

此滤波器可以设置 $F(0,0)$ 为零(正如所希望的那样),而保留其他傅里叶变换的频率成分不变。处理后的图像(具有 0 平均值)可以通过对 $H(u,v)F(u,v)$ 进行傅里叶反变换来获得,如式(4.2.28)所示。如前所述,$F(u,v)$ 的实部和虚部都乘以滤波函数 $H(u,v)$。

刚刚讨论的滤波器称为"陷波滤波器",因为它除了原点处有凹陷外其他均是常量函数。图 4.4(a)中的图像经此滤波器处理的结果显示在图 4.6 中。注意,由于使平均值设置为 0 而产生整体平均灰度级的降低;同时还要注意产生边缘突出的伴随结果[实际中,所显示的图像平均值不能为零,因为如果其平均灰度级为零,图像必为负值,而显示无法处理负值。图 4.6 是用"标准"方式显示的,将所有负值当做 0 显示(即黑色),而其他值在此基础上按比例增加]。如 5.4.3 节所示,当可以识别由特定的、局部化频域成分引起的空间图像效果时,陷波滤波器是一个非常有用的工具。

在傅里叶变换中,低频主要决定图像在平滑区域中总体灰度级的显示,而高频决定图像细节部分,如边缘和噪声。这些思想将在下一节中详细讨论,但用其他两类滤波器的例子有益于补充对陷波滤波器的说明。使低频通过而使高频衰减的滤波器称为"低通滤波器"。具有相反特性的滤波器称为"高通滤波器"。可以预料,被低通滤波的图像比原始图像少一些尖锐的细节部分,因为高频部分已被衰减。同样,被高通滤波的图像在平滑区域中将减少一些灰度级的变化并突出过渡(如边缘)灰度级的细节部分。这样的图像将更为锐化。

图 4.7 说明了低通和高通滤波对图 4.4(a)中图像的影响。图的左侧部分显示了滤波器,右侧部分显示了使用图 4.5 总结出的过程进行滤波的结果。而滤波函数 $H(u,v)$ 是周期对称

的。在将原点移至由 $F(u,v)$ 设置的矩形区域中心后，它们乘以中心变换，正如式(4.2.27)，式(4.2.28)和图4.5指出的那样。取每个结果的实部，然后乘以 $(-1)^{x+y}$ 以得到图4.7右侧部分所显示的图像。正如所料，图4.7(b)中的图像被模糊了，而在图4.7(d)中的图像是锐化的，因为 $F(0,0)$ 已被设置为0，所以几乎没有平滑的灰度级细节。这是高通滤波的典型结果，接下来的过程通常在滤波器中加入常量，以使 $F(0,0)$ 不被完全消除。使用此过程的结果如图4.8所示。它对图4.7的改进很明显。

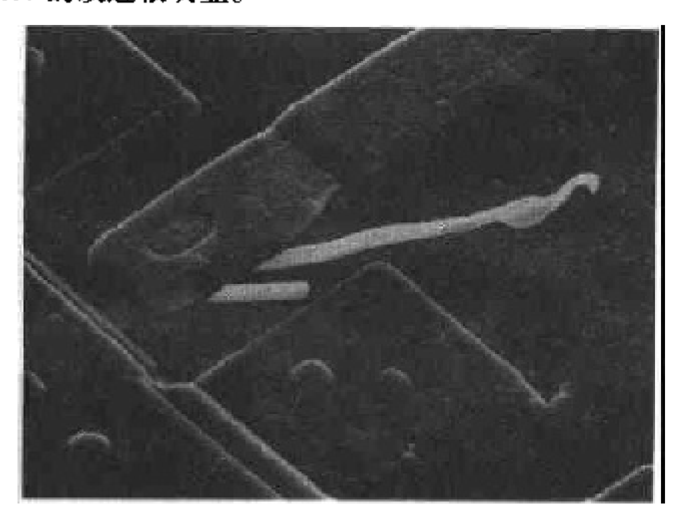


图4.6　图4.4(a)中的图像经陷波滤波器处理后的结果，在傅里叶变换中，陷波滤波器将原点设置为0

4.2.4　空间域滤波和频率域滤波之间的对应关系

在前一章中通过观察和(或)使用数学公式得到了不同的空间滤波器形式，如拉普拉斯算子。在本节中，将建立空间域中的一些滤波器和其频率域对应项之间的直接联系。

空间域和频率域之间最基本的联系是由卷积定理的著名结论建立的。读者已在空间域熟悉了卷积的基本概念和机理(在3.5节中已说明过)。将图像的模板在图像中逐像素移动，并对每个像素进行指定数量的计算的过程就是卷积过程。形式上，大小为 $M\times N$ 的两个函数 $f(x,y)$ 和 $h(x,y)$ 的离散卷积表示为 $f(x,y)*h(x,y)$，并定义如下：

$$f(x,y)*h(x,y)=\frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}f(m,n)h(x-m,y-n) \tag{4.2.30}$$

除了前面的常数、负号以及求和的上下限之外，这个表达式与式(3.5.1)相似。特别是负号只说明函数 h 关于原点镜像对称。这是卷积定义中自带的。式(4.2.30)只不过是一种实现：(1)关于原点翻转函数；(2)通过改变 (x,y) 的值相对于一个函数移动另外一个函数；(3)对每一个 (x,y) 的位移值，计算所有 m 和 n 值乘积的和。(x,y) 位移是以整数增加的，当函数不再重叠时停止。

用 $F(u,v)$ 和 $H(u,v)$ 分别表示 $f(x,y)$ 和 $h(x,y)$ 的傅里叶变换，卷积定理的一部分只说明 $f(x,y)*h(x,y)$ 和 $F(u,v)H(u,v)$ 组成傅里叶变换对。形式上表示如下：

$$f(x,y)*h(x,y)\Leftrightarrow F(u,v)H(u,v) \tag{4.2.31}$$

双箭头表示左边的表达式(空间域卷积)可以通过对右边表达式[在频率域的 $F(u,v)H(u,v)$ 的乘积]进行反傅里叶变换获得。相反，右边表达式可以通过对左边表达式进行正向傅里叶变换获得。类似的结果是频率域的卷积被简化为空间域的乘法，反之亦然；即：

$$f(x,y)h(x,y)\Leftrightarrow F(u,v)*H(u,v) \tag{4.2.32}$$

这两个结论构成卷积定理。如上所述的定理并没有复杂之处,卷积操作过程就是如此。处理过程的其他部分仅为两个函数逐元素相乘的乘法。

图 4.7 (a)一个二维低通滤波函数,(b)图 4.4 中图像经低通滤波的结果,(c)二维高通滤波函数,(d)图4.4中图像经高通滤波的结果

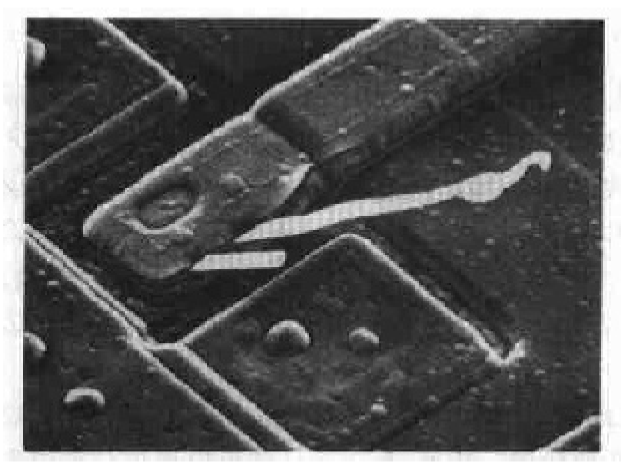


图 4.8 用图 4.7(c)中的滤波器对图 4.4(a)中的图像进行高通滤波的结果,该滤波器通过对滤波函数加上一个滤波器高度一半的常数加以改进

在实现空间域和频率域之间的联系前,还需要一个概念。在(x_0,y_0)坐标处强度为 A 的冲激函数表示为$A\delta(x-x_0,y-y_0)$,并定义如下:

$$\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}s(x,y)A\delta(x-x_0,y-y_0)=As(x_0,y_0) \tag{4.2.33}$$

总之,这个等式表明乘以一个冲激的函数 $s(x,y)$的和等于冲激函数在此位置的值再乘以冲激强度 A。函数的范围限制即为求和的上下限。将 $A\delta(x-x_0,y-y_0)$指定为一幅尺寸为$M\times N$的图像。它只在(x_0,y_0)处有为 A 的图像值,其他处的值全为零。

通过设式(4.2.30)中的f或h 为冲激函数,并使用式(4.2.33)中的定义稍加处理就可以得出带冲激函数的卷积"复制"了在冲激位置上此函数的值。此特性称为卷积函数的"筛选"特性。此时,特别重要的是在原点处的单位冲激情况,用 $\delta(x,y)$表示。这里,

$$\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}s(x,y)\delta(x,y)=s(0,0) \tag{4.2.34}$$

利用这个简单的工具,现在可以在空间域和频率域滤波之间建立更为有趣和有用的联系了。根据式(4.2.16),可以计算原点处单位冲激的傅里叶变换:

$$\begin{aligned}F(u,v)&=\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\delta(x,y)e^{-j2\pi(ux/M+vy/N)}\\&=\frac{1}{MN}\end{aligned} \tag{4.2.35}$$

其中,第二步根据式(4.2.34)而来。由此,可以得出空间域原点处的冲激函数的傅里叶变换是实常量(这意味着相角为0)。如果冲激任意放置,变换会含有更为复杂的成分。它们的幅度应该相同,但在变换后出现的非零相角会引起脉冲平移。

现在假设$f(x,y)=\delta(x,y)$,同时执行式(4.2.30)中定义的卷积,再次使用式(4.2.34)得到:

$$\begin{aligned}f(x,y)*h(x,y)&=\frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{n-1}\delta(m,n)h(x-m,y-n)\\&=\frac{1}{MN}h(x,y)\end{aligned} \tag{4.2.36}$$

其中,注意在求和时的变量为 m 和n,最后一步由式(4.2.34)而来。将式(4.2.35)、式(4.2.36)和式(4.2.31)合并,得到:

$$\begin{gathered}f(x,y)*h(x,y)\Leftrightarrow F(u,v)H(u,v)\\\delta(x,y)*h(x,y)\Leftrightarrow\Im[\delta(x,y)]H(u,v)\\h(x,y)\Leftrightarrow H(u,v)\end{gathered} \tag{4.2.37}$$

仅使用冲激函数和卷积定理的性质,确定了在空间域和频率域中的滤波器组成了傅里叶变换对。因此,给出在频率域的滤波器,可以通过将前者进行反傅里叶变换而得到在空间域相应的滤波器,反之亦然。

注意，前述的所有函数均为相同的尺寸 $M\times N$ 。因此，在实际中，指定一个频率域的滤波器，然后进行反变换以计算相同尺寸的空间域的相应滤波器，但这种方法从计算的角度来看并不能解决太大问题。如 4.6 节讨论的，如果两个滤波器是相同尺寸，那么通常在频率

域进行滤波计算更为有效。但是，在空间域更适用于更小的滤波器。这正是我们所感兴趣的联系。滤波在频率域中更为直观，但在空间域使用更小的滤波器模板更为明智。式(4.2.37)告诉我们，可以在频率域指定滤波器，做反变换，然后在空间域使用结果滤波器作为在空间域构建更小空间滤波模板的指导(更为正式的方法在4.6.7节中论)。这些在以下例证加以说明。记住，在以下的讨论中傅里叶变换和反变换都是线性过程(习题4.2)，所以讨论仅限在线性滤波中。

基于高斯函数的滤波有特殊的重要性，因为它们的形状易于确定，而且高斯函数的傅里叶变换和反变换均为实高斯函数。在此将讨论限制在一个变量的范围以简化符号表示。二维函数将在本章的后面讨论。

用 $H(u)$表示频率域，高斯滤波器函数由如下等式给出：

$$H(u) = Ae^{-u^2/2\sigma^2} \tag{4.2.38}$$

其中，σ 为高斯曲线的标准差。可以看出相关的空间域滤波器为(习题4.4)：

$$h(x) = \sqrt{2\pi}\sigma Ae^{-2\pi^2\sigma^2x^2} \tag{4.2.39}$$

这两个等式表示了一个重要结论，原因有二：(1)它们组成了傅里叶变换对，成分均为实高斯函数，这非常有助于分析，因为我们不必处理复数量，除此之外，高斯曲线直观且易于操作；(2)这些函数有相互之间的作用，换句话说，当 $H(u)$有很宽的轮廓时(大的 σ 值)，$h(x)$有很窄的轮廓，反之亦然。事实上，当 σ 接近无限时，$H(u)$趋于常量函数，而 $h(x)$趋于冲激函数。这正是在4.2节图4.2和图4.3中看到的相互作用类型的特性。这两个特性有助于理解空间域和频率域之间的滤波性质，因为它们有相似的分析解释。

高斯滤波器在频率域的曲线如图4.9(a)所示。读者可以识别出 $H(u)$的形状为低通滤波器。相应的空间域低通滤波器如图4.9(c)所示。我们的兴趣在于 $h(x)$的一般形状，通常希望作为指导来确定空间域中较小滤波器的系数。两个滤波器的相似之处在于两个域中的值均为正。因此得出结论，在空间域通过使用带正系数的模板可以实现低通滤波，做法如3.6.1节所示。该节中的两种模板作为参考如图4.9(c)所示。另一个重要特性是在前一段中讨论的相互之间的关系。频率域滤波器越窄，滤除的低频成分就越多，使得图像就越模糊。在空间域，这意味着滤波器就越宽，模板就越大，如例3.9所示。

更复杂的滤波器可以由式(4.2.38)的基本高斯函数构造出来。例如，可以构造一个不同的高斯高通滤波器，如下所示：

$$H(u) = Ae^{-u^2/2\sigma_1^2} - Be^{-u^2/2\sigma_2^2} \tag{4.2.40}$$

其中，$A \geqslant B$，并且 $\sigma_1 > \sigma_2$。在空间域的相应滤波器为：

$$h(x) = \sqrt{2\pi}\sigma_1 Ae^{-2\pi^2\sigma_1^2x^2} - \sqrt{2\pi}\sigma_2 Be^{-2\pi^2\sigma_2^2x^2} \tag{4.2.41}$$

这两个函数的曲线分别示于图4.9(b)和(d)中。还要注意宽度上的相互关系，但其中最重要的特性是空间域滤波器有正值和负值。事实上，应特别注意一旦值变为负数，就再也不会变为正数。在第3章中高通滤波使用的两个模板如图4.9(d)所示。空间曲线和滤波器之间形式上的相似很明显。

在第3章中，我们指出低通和高通滤波器的形状严格基于空间域的考虑。通过在频率域的分析已得出所有小的空间域滤波器模板的基本形状，如图4.9所示。虽然付出了巨大的努

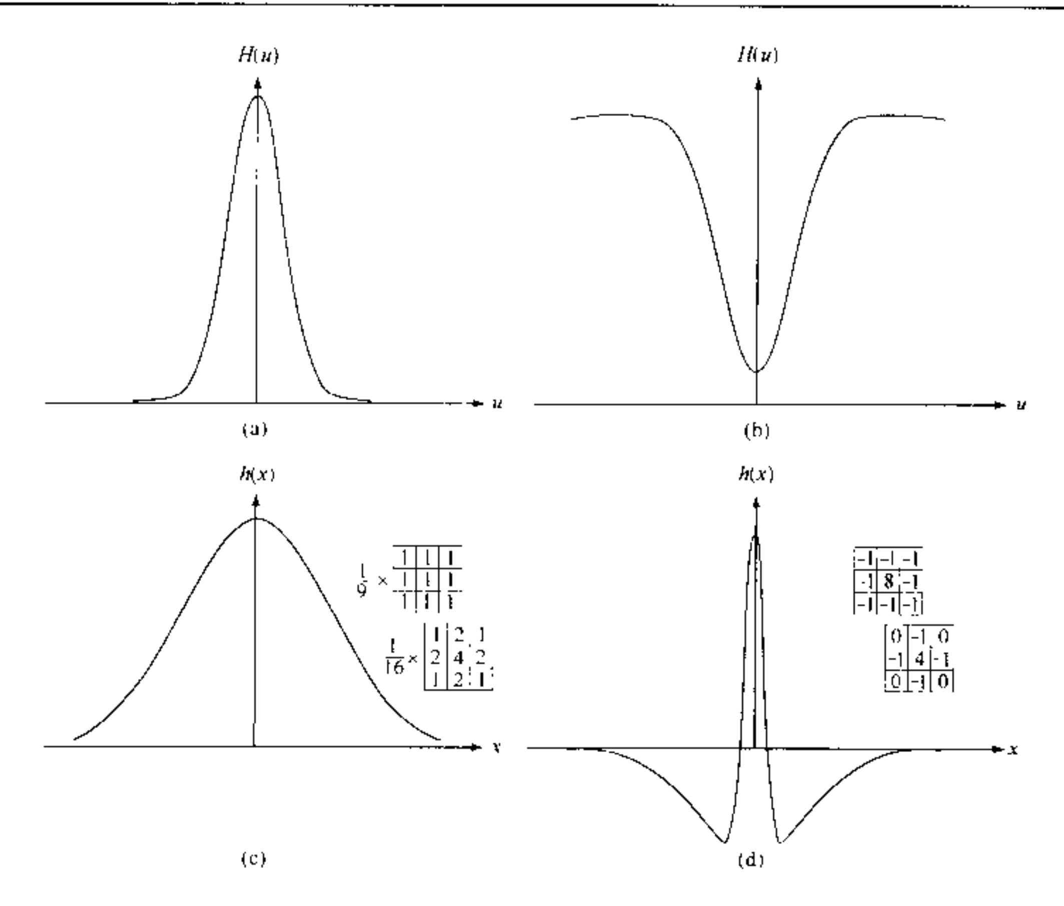


图 4.9　(a)频率域高斯低通滤波器,(b)频率域高斯高通滤波器,(c)相应的空间低通滤波器,(d)相应的空间高通滤波器。所示的模板是第3章中低通和高通滤波器采用的

力才得到此结论,但如果没有已建立的基础要真正理解频率域滤波是很困难的。

在频率域技术的发展过程中,总会出现的问题是计算的复杂性。为什么在频率域中的内容(至少是部分的)在空间域使用小的空间模板?基本的回答是双方面的。首先,正如我们看到的,频率域在很大程度上凭直观指定滤波器;另一方面取决于空间模板的大小,通常根据可比较的方法来决定。

通常基于此目的所使用的基准是空间域和频率域中应用的卷积。空间域的卷积在式(4.2.30)中给出,而且从卷积定理得出:通过两个函数变换结果在频率域的反变换可以得到相同的结论。假定在同一台微机上用软件实现这两种方法[用在 4.6.6 节讨论的快速傅里叶变换(FFT)算法进行频率域的计算]。会发现在频率域因为 M 和 N 的值较小所以执行速度非常快。例如,Brigham [1988]的比较显示,对于一维情况,如果点数大于 32,则 FFT 的方法会更快。虽然这个数字部分决定于其他因素,如所使用的机器和算法,但它通常低于在图像处理中遇到的值。

频率域可以看成一个“实验室”,可以从中利用频率成分和图像外表之间的对应关系。如在本章后面多次证明的,一些在空间域直接表述非常困难的、甚至是不可能的增强任务在频率域中变得非常普通。一旦通过频率域实验选择了空间滤波,通常实际实施都在空间域进行。一种方法是指定小的空间模板以试图捕捉空间域内完整滤波器函数的“本质”,如图 4.9中所解释的那样。更为普通的方法是利用基于数学和统计准则的近似设计一个二维数字滤波器。

在4.6.7节中还会谈到这一点。

4.3 平滑的频率域滤波器

如4.2.3节指出的,边缘和其他尖锐变化(如噪声)在图像的灰度级中主要处于傅里叶变换的高频部分。因此,平滑(模糊)可以通过衰减指定图像傅里叶变换中高频成分的范围来实现。

在频率域中,基本的滤波"模型"由式(4.2.27)给出,在这里为方便起见再次给出:

$$G(u,v) = H(u,v)F(u,v) \tag{4.3.1}$$

其中,$F(u,v)$是被平滑的图像傅里叶变换。目标是选择一个滤波器变换函数 $H(u,v)$,以通过衰减 $F(u,v)$的高频成分产生 $G(u,v)$。所有在本节中进行的滤波都基于4.2.3节中指出的过程,包括使用"零相移"的滤波器。

这里考虑三种滤波器:理想滤波器、巴特沃思滤波器和高斯滤波器。这三种滤波器涵盖了从非常尖锐(理想)到非常平坦(高斯)范围的滤波器函数。巴特沃思滤波器有一个参数,称为滤波器的"阶数"。当此参数的值较高时,巴特沃思滤波器接近理想滤波器。因此,巴特沃思滤波器可看做两种"极端"滤波器的过渡。

4.3.1 理想低通滤波器

所想像的最简单的低通滤波器是"截断"傅里叶变换中所有高频成分,这些成分处在距变换原点的距离比指定距离 D_0 要远得多的位置。这种滤波器称为二维理想低通滤波器(ILPF),其变换函数为:

$$H(u,v) = \begin{cases} 1 & D(u,v) \leqslant D_0 \\ 0 & D(u,v) > D_0 \end{cases} \tag{4.3.2}$$

其中,D_0 是指定的非负数值,$D(u,v)$是(u,v)点距频率矩形原点的距离。如果要研究的图像尺寸为 $M \times N$,它的变换也有相同的尺寸,所以如式(4.2.21)中讨论的那样,由于变换被中心化了,所以,频率矩形的中心在$(u,v)=(M/2,N/2)$处。在这种情况下,从点(u,v)到傅里叶变换中心(原点)的距离如下所示:

$$D(u,v) = \left[(u - M/2)^2 + (v - N/2)^2\right]^{1/2} \tag{4.3.3}$$

图4.10(a)显示了 $H(u,v)$作为 u 和 v 函数的三维透视曲线,图4.10(b)将 $H(u,v)$作为图像显示。"理想滤波器"的名称表明在半径为 D_0 的圆内,所有频率没有衰减地通过滤波器,而在此半径的圆之外的所有频率完全被衰减掉。在本章所考虑的滤波器是关于原点辐射状对称的。这意味着从原点沿着半径线延伸的距离函数的横截面足以满足一个指定的滤波器,如图4.10(c)所示。完整的滤波器变换函数可将此横截面绕着原点旋转360°来实现。

对于理想的低通滤波器横截面,在 $H(u,v)=1$ 和 $H(u,v)=0$ 之间的过渡点称为"截止频率"。例如,在图4.10中,其截止频率为 D_0。理想低通滤波器的这种陡峭的截止频率是不能用电子部件实现的,尽管它们可以在计算机上实现。在数字图像中使用这些"非物理可实现"的滤波器的作用将在本节后面的部分中讨论。

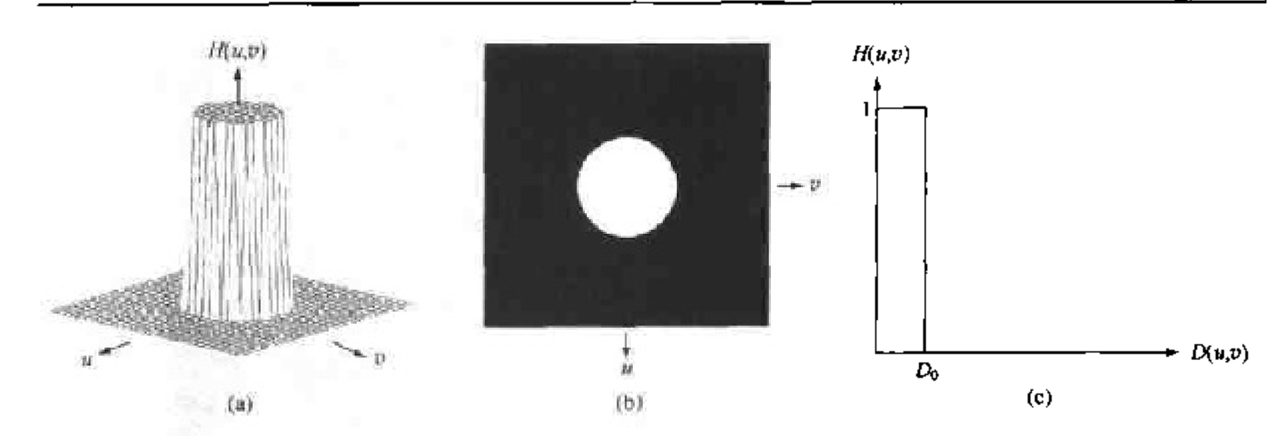


图 4.10　(a)理想低通滤波器变换函数的透视图,(b)以图像显示的滤波器,(c)滤波器的径向横断面

在本节中介绍的低通滤波器用具有相同截止频率的函数研究它们的特性而加以比较。一种建立一组标准截止频率位置的方法是计算包含指定总图像功率值 P_T 的范围。这个值是将每个点(u,v)的功率谱成分相加得到的,其中 $u=0,1,2,\cdots,M-1$,$v=0,1,2,\cdots,N-1$,即:

$$P_T = \sum_{u=0}^{M-1}\sum_{v=0}^{N-1} P(u,v) \tag{4.3.4}$$

其中,$P(u,v)$ 在式(4.2.20)中给出。如果变换被中心化、原点在频率矩形的中心、半径为 r 的圆包含 $\alpha\%$ 的功率,其中

$$\alpha = 100\left[\sum_u \sum_v P(u,v)/P_T\right] \tag{4.3.5}$$

总和取处于圆之内或边界线上的(u,v)值。

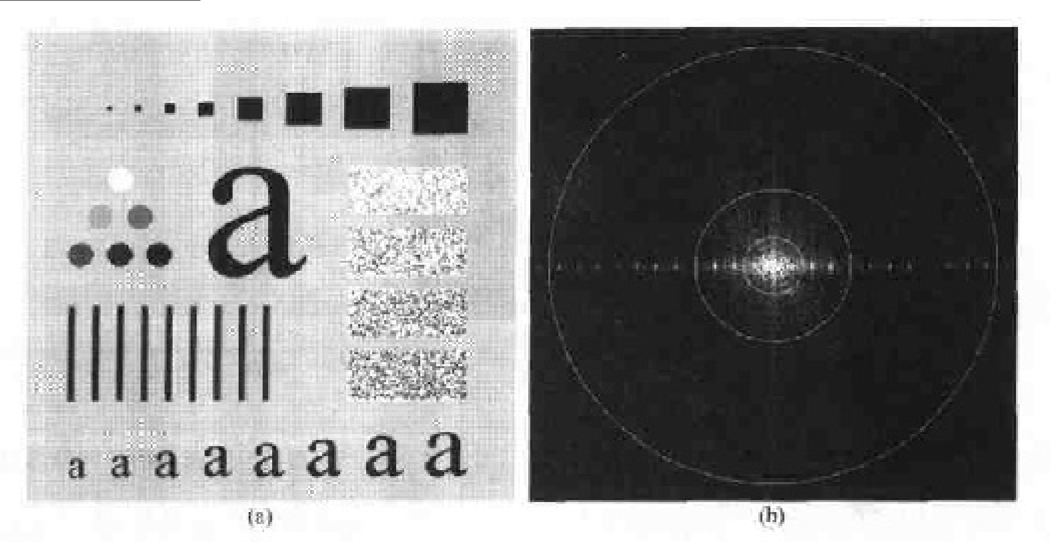
例 4.4　作为距 DFT 原点距离函数的图像功率

图 4.11(a)显示了在图 3.35 中用来说明空间模糊的测试模式。这幅图像的傅里叶谱示于图 4.11(b)。在谱中叠加的圆周分别有 5,15,30,80 和 230 像素的半径(半径为 5 的圆周不易看见)。这些圆周包围的图像功率的百分比分别为 92.0%,94.6%,96.4%,98% 和 99.5%。谱迅速衰落,92% 的功率包含在相对较小的半径为 5 的圆周之内。

图 4.12 显示了应用图 4.11(b)所示半径处截止频率的理想低通滤波器的结果。图 4.12(b)没有实际意义,除非在这种情况下模糊的目标是要消除图像的所有细节,只用“斑点”来代表最大的物体。在此图像中严重的模糊表明,图像中多数尖锐的细节信息包含在被滤除掉的 8% 的功率之内。随着滤波器半径的增加,越来越少的功率被滤除掉,使得模糊减弱。注意,在从图 4.12(c)到(e)中都有“振铃”特征,而当被滤除的高频内容的数量减少时,图像的质量逐渐变好。振铃是明显的,因为图像中只有 2% 的功率被除掉了。这种振铃现象是理想滤波器的特性,稍后将进行解释。最后,当 $\alpha=99.5\%$ 时的观察结果在噪声区域内有非常小的模糊,大部分图像与原始图像非常接近。这表明在此特殊情况下,图像的边缘信息包含在 0.5% 以上的谱功率中。

从这个例子中,可以清楚地看到理想低通滤波器非常不实用。然而,因为理想滤波器可以在计算机上实现,所以作为滤波概念发展的一部分,研究滤波器的特性非常有用。同时,如以下讨论所示,在试图解释 ILPF 在空间域的振铃特性时可以得到一些有趣的现象。

ILPF 的模糊和振铃特性可参考如 4.2.4 节讨论的卷积定理来解释。原始图像 $f(x,y)$的



(a) (b)

图 4.11 (a)尺寸为 500×500 像素的图像,(b)图像的傅里叶谱。叠加的圆环具有 5,15,30,80 和 230 像素的半径,它们分别包含了 92.0%,94.6%,96.4%,98.0% 和 99.5%的图像功率

傅里叶变换和模糊的图像 $g(x,y)$在频率域中相互关联,如下式所示:

$$G(u,v) = H(u,v)F(u,v)$$

其中,与前面一样,$H(u,v)$还表示滤波器函数,而 F 和 G 分别是两个刚刚提到的图像的傅里叶变换。卷积定理指出在空间域的相应过程:

$$g(x,y) = h(x,y) * f(x,y)$$

其中 $h(x,y)$是滤波器变换函数 $H(u,v)$的反变换。

在空间域将模糊作为卷积过程来理解的关键是 $h(x,y)$的特性。例如,在前述例子中产生模糊且半径为 5 的 ILPF 如图 4.13(a)所示。这是在频率域的函数 $H(u,v)$。空间滤波器函数 $h(x,y)$可以通过标准方法获得:(1)使 $H(u,v)$乘以$(-1)^{u+v}$,从而移到中心;(2)然后紧接着做反傅里叶变换;(3)反 DFT 的实部乘以$(-1)^{x+y}$。图 4.13(b)显示了此过程的结果。

我们看到滤波器 $h(x,y)$有两个主要特性:在原点处的一个主要成分,及中心成分周围集中、呈周期性的成分。中心成分主要决定模糊。集中的成分主要决定了理想滤波器振铃现象的特性。中心成分的半径和距原点每单位距离上周期的数量都与理想滤波器的截止频率值成反比。在顶部的插图是通过空间滤波器中心的水平扫描线的灰度级剖面线。所示的轴表明了零幅度,所以,我们看到空间域滤波器有负值。通常这不是严重的问题,因为较大的中心成分决定着卷积的结果。然而,被滤波的图像能够有负值,因此通常要求做标定。

下面,假定 $f(x,y)$是一幅简单图像,它由在黑色背景下五个明亮的像素组成,如图 4.13(c)所示。这些明亮的点可被近似地看做冲激,其强度决定于点的亮度。然后,$h(x,y)$和 $f(x,y)$的卷积仅仅是在每个冲激处“复制”$h(x,y)$的过程,如 4.2.4 节中所述。此操作的结果,如图 4.13(d)所示,解释了原始点如何通过 $f(x,y)$和模糊滤波器函数 $h(x,y)$的卷积而变模糊。同样注意在此过程中对振铃的介绍。事实上,振铃在此种情况下很严重,以至于由相互之间的干扰而产生畸变。这些概念被扩展到更复杂的图像上,这里,考虑将每个像素作为一个脉冲,而且其强度与像素的灰度级成比例。图 4.13 底部的插图显示了通过被滤波图像中心的对角扫描线的灰度级剖面线。

$H(u,v)$和 $h(x,y)$之间的互易性及刚刚讨论的卷积过程,在数学上解释了为什么模糊和

振铃比频率域上的窄带滤波更为严重。这种互易特性应成为读者惯用的方式。在后面两节中将说明有可能出现很少振铃或没有振铃的模糊，这是我们主要的目标。

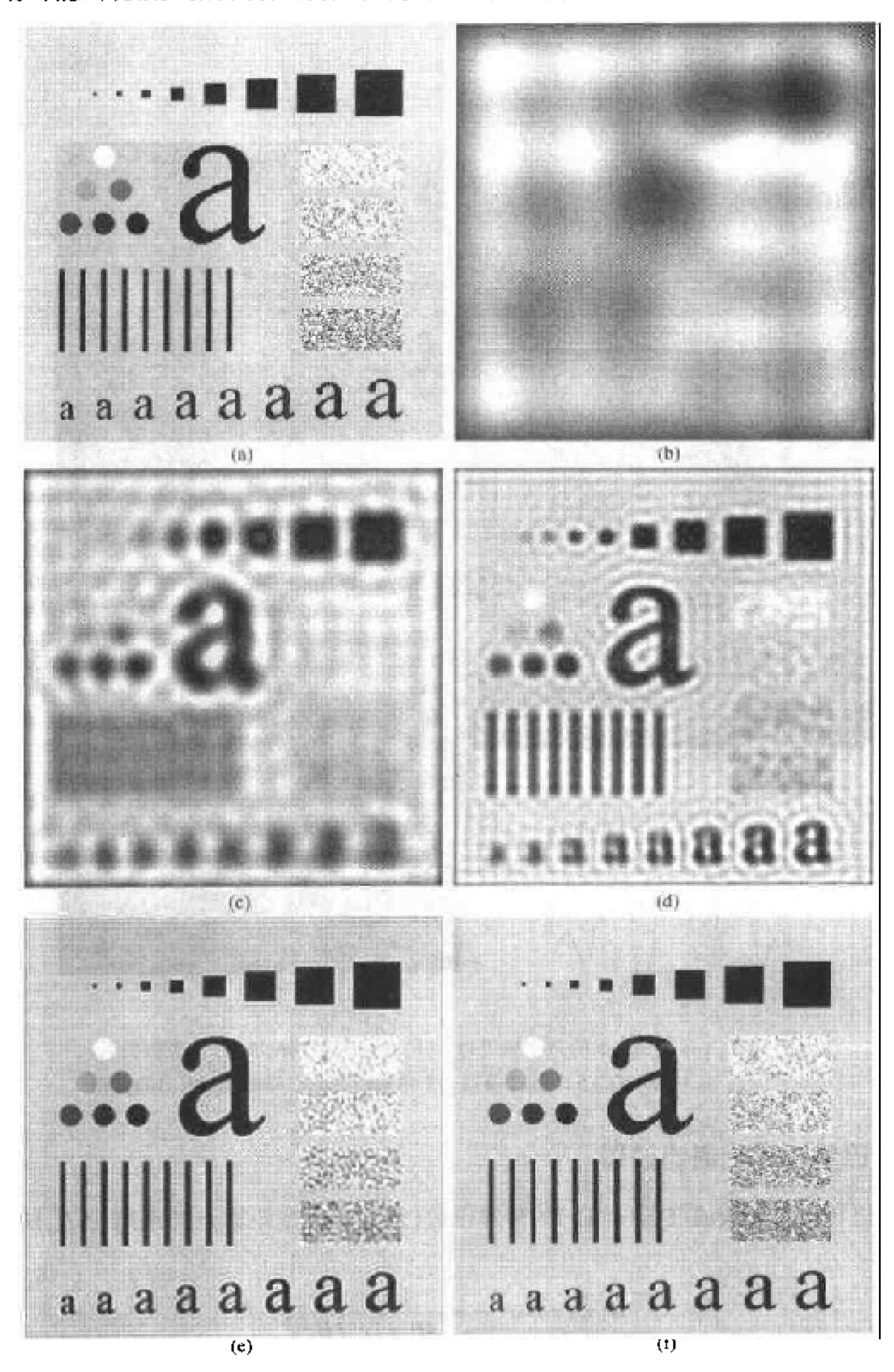


图 4.12　(a)原图像，(b)～(f)用如图 4.11 所示的半径值为 5,15,30,80 和 230 的截止频率进行理想低通滤波的结果。由这些滤波器滤除的功率分别为总功率的 8%，5.4%，3.6%，2%和 0.5%

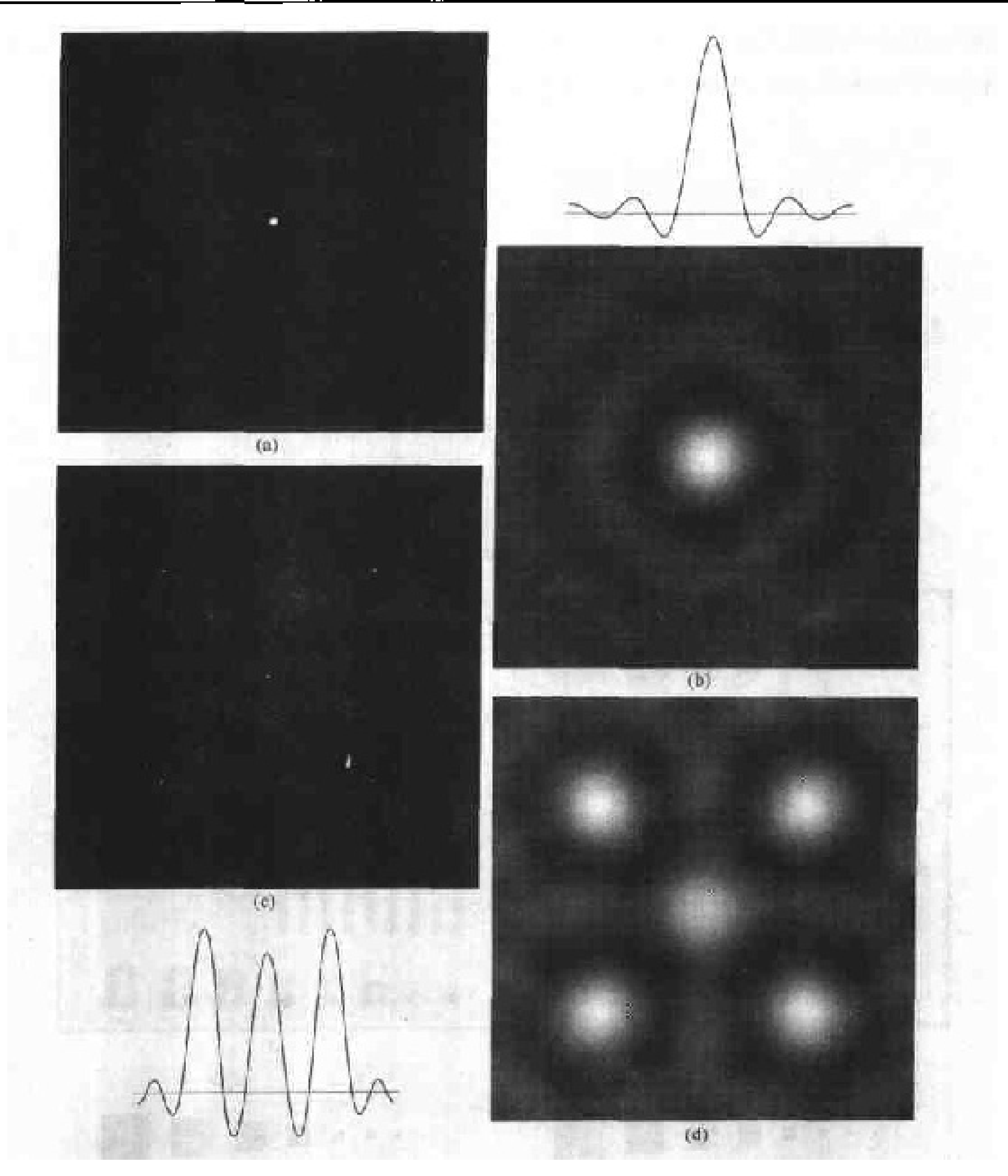


图 4.13 (a)半径为 5 的频率域 ILPF,(b)相应的空间滤波器(注意振铃),(c)空间域的 5 个脉冲模拟 5 个像素值,(d)空间域(b)和(c)的卷积

4.3.2 巴特沃思低通滤波器

n 级巴特沃思低通滤波器(BLPF)的传递函数(且截止频率距原点的距离为 D_0)的定义如下:

$$H(u,v)=\frac{1}{1+[D(u,v)/D_0]^{2n}} \tag{4.3.6}$$

其中,$D(u,v)$由式(4.3.3)给出。其透视图、图像显示以及 BLPF 函数的径向横截面如图 4.14 所示。

不同于 ILPF,BLPF 变换函数在通带与被滤除的频率之间没有明显的截断。对于有平滑传

递函数的滤波器，定义一个截止频率的位置并使 $H(u,v)$ 幅度降到其最大值的一部分。在式(4.3.6)中，当 $D(u,v)=D_0$ 时，$H(u,v)=0.5$(从最大值 1 降到它的 50%)。

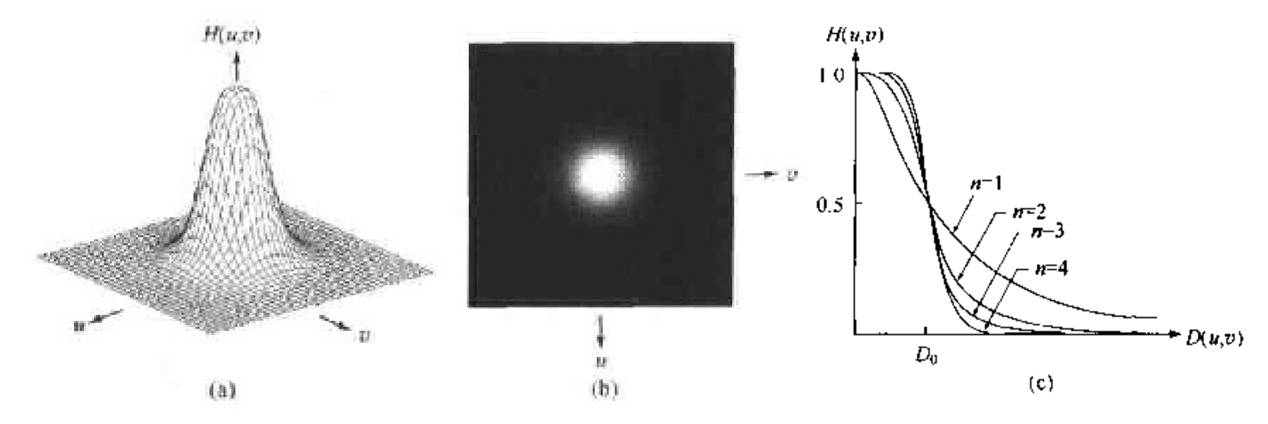


图 4.14　(a)巴特沃思低通滤波器函数的透视图，(b)以图像显示的滤波器，(c)阶数从 1 到 4 的滤波器横截面

例 4.5　巴特沃思低通滤波器

图 4.15 显示了将式(4.3.6)中的 BLPF 应用于图 4.15(a)中的结果，其中 $n=2$ 并且 D_0 与图 4.11(b)所示的 5 个半径相等。与图 4.12 中所示的 ILPF 结果不同，这里注意到在模糊处理中作为增加的截止频率函数的平滑传递函数。而且，用特殊的 BLPF 在任何图像处理过程中都不会出现振铃，这取决于低频和高频之间的滤波器的平滑过渡。

一个一阶的巴特沃思滤波器没有振铃。在二阶中振铃通常很微小，但阶数增高时振铃便成为一个重要因素。图 4.16 显示了不同阶数(具有 5 个像素的截止频率)的 BLPF 在空间表示上的有趣对比。同时显示了通过每个滤波器中心的水平扫描线的灰度级剖面线。这些滤波器是通过使用与图4.13(b)相同的过程获得并显示的。为了便于比较，用伽马变换[参见式(3.2.3)]对图 4.16 中的图像进行了额外增强，以突出远离原点的成分。阶数为 1 的 BLPF [图 4.16(a)]没有振铃和负值。二阶的 BLPF 显示了轻微的振铃和较小的负值，但它们远没有 ILPF 明显。如其余图像显示的那样，在 BLPF 中，振铃对高阶滤波器很明显。一个 20 阶的巴特沃思滤波器已经呈现出 ILPF 的特性，这可以从图 4.16(d)和图 4.13(b)的比较中看出。在此限制下，两种滤波器完全相同。总之，二阶的 BLPF 是在有效的低通滤波和可接受的振铃特性之间的折中。

4.3.3　高斯低通滤波器

一维高斯低通滤波器(GLPF)在 4.2.4 节中已介绍过，用来帮助寻找空间域与频率域之间的重要联系。这些滤波器的二维形式由下式给出：

$$H(u,v)=e^{-D^2(u,v)/2\sigma^2} \tag{4.3.7}$$

其中，如在式(4.3.3)中那样，$D(u,v)$是距傅里叶变换原点的距离，假定已使用 4.2.3 节中的过程将变换移至频率区域的中心。我们不使用 4.2.4 节中的滤波器前的常数，以使所有本节讨论的其他滤波器达到一致(即在原点处为 1)。如以前一样，σ 表示高斯曲线扩展的程度。使 $\sigma=D_0$，可以将滤波器表示为本节标记法下更为熟悉的形式：

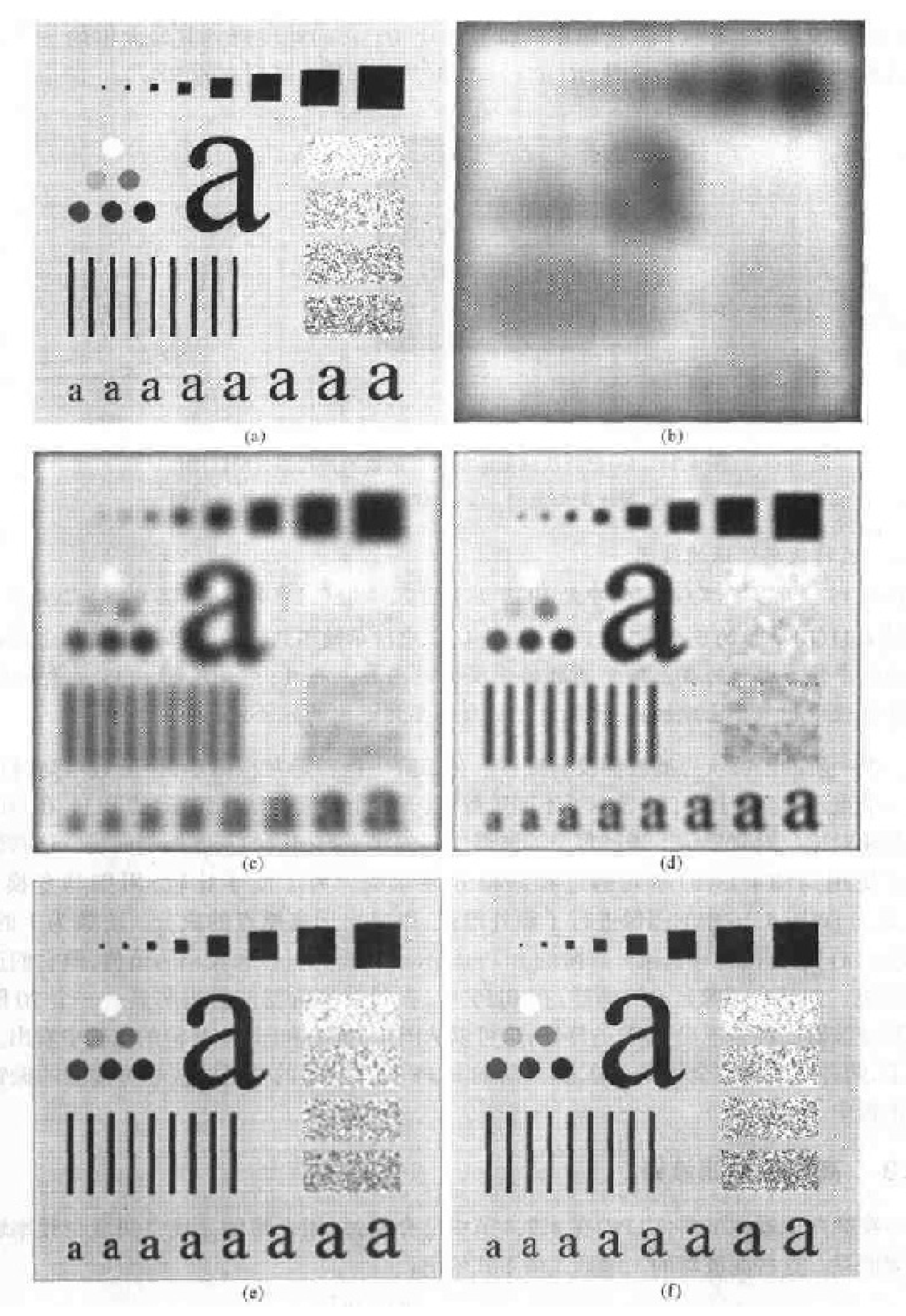


图 4.15 (a)原图像,(b)~(f)用二阶 BLPF 滤波的结果,如图 4.11(b)所示,截止频率点半径为 5,15,30,80 和 230。可与图 4.12 做比较

$$H(u,v) = e^{-D^2(u,v)/2D_0^2} \tag{4.3.8}$$

其中，D_0 是截止频率。当 $D(u,v)=D_0$ 时，滤波器下降到它最大值的 0.607 处。

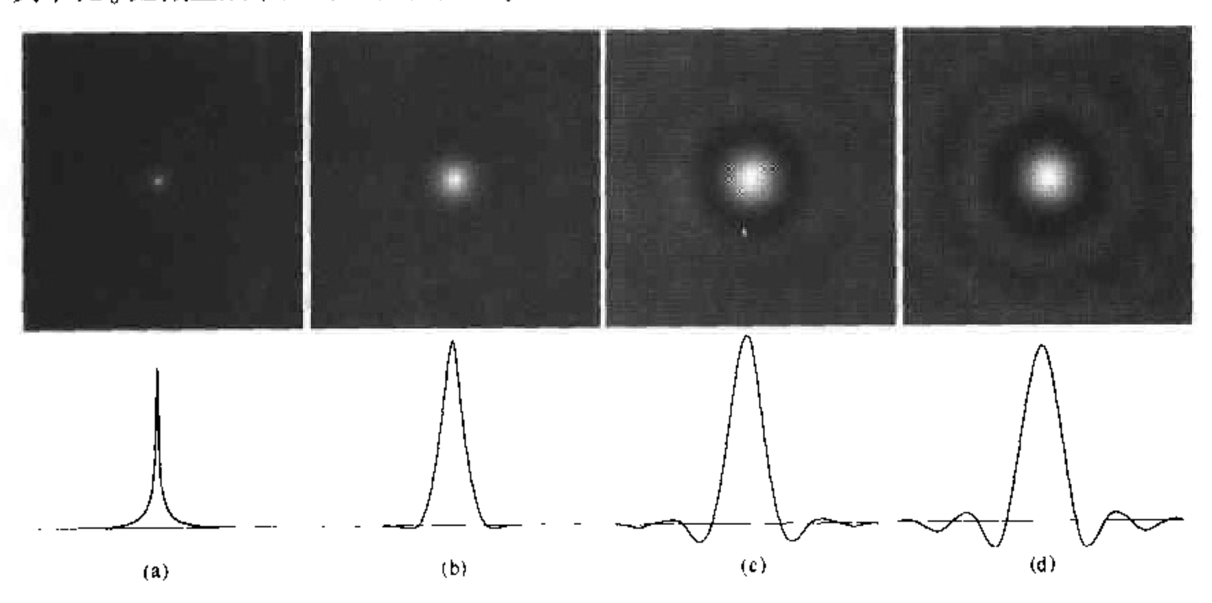


图 4.16　(a)～(d)阶数为 1,2,5 和 20 的 BLPF 的空间表示及相应的通过滤波器中心的灰度级剖面图(所有的滤波器都有半径为 5 的截止频率)。注意振铃作为滤波器阶数的函数越发明显

如 4.2.4 节中讨论的，高斯低通滤波器的傅里叶反变换也是高斯的。我们已经看到作为分析工具这种函数特性的优点了。这也意味着通过计算式(4.3.7)或式(4.3.8)的傅里叶反变换而得到的空间高斯滤波器将没有振铃。其透视图、图像显示以及 GLPF 函数的径向横截面如图 4.17 所示。

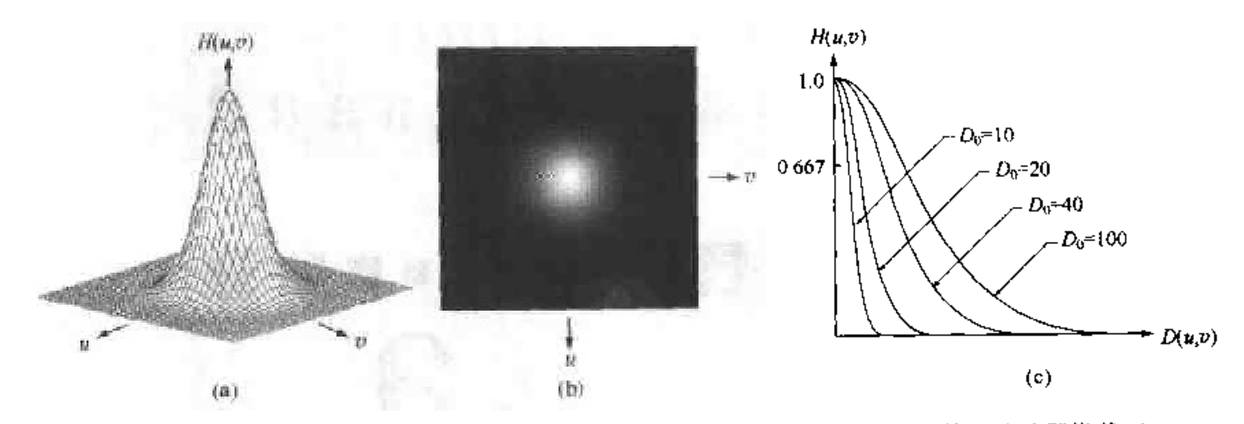


图 4.17　(a)GLPF 传递函数的透视图，(b)以图像显示的滤波器，(c)各种 D_0 值的滤波器横截面

例 4.6　高斯低通滤波器

图 4.18 显示了将式(4.3.8)的 GLPF 应用到图 4.18(a)的结果，并且 D_0 等于图 4.11(b)所示的 5 个半径。如 BLPF(图 4.15)的情况那样，注意到作为增大截止频率的函数在模糊中的平滑变换。GLPF 不能达到有相同截止频率的二阶 BLPF 的平滑效果，如比较图 4.15(c)和图 4.18(c)时我们所看到的。这正如我们所料，因为 GLPF 的剖面线没有二阶 BLPF 的剖面线紧凑。然而，结果可在总体上相比较，我们确信在 GLPF 中没有振铃。这是实际中

的一个重要特性,尤其是在人工图像(如医学图像)不太能接受的情况下。在需要严格控制低频和高频之间截止频率的过渡的情况下,BLPF 为更合适的选择。额外控制滤波器剖面线的代价是可能产生振铃。

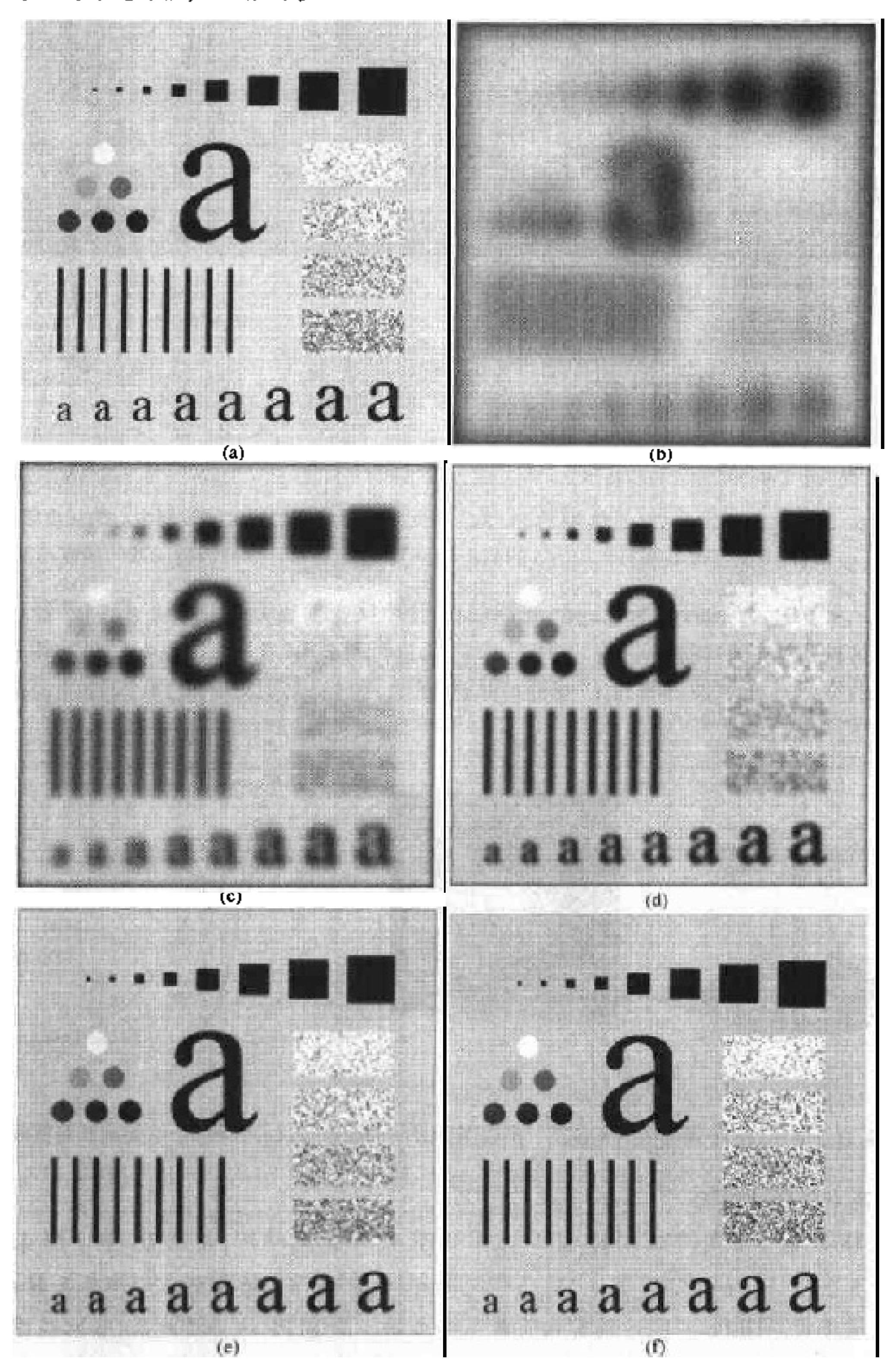


图 4.18 (a)原图像,(b)~(f)用高斯低通滤波器滤波的结果,其截止频率如图 4.11(b)所示,一组半径值为5,15,30,80和230。可比较图4.12和图4.15

4.3.4　低通滤波的其他例子

为了说明和比较滤波效果，迄今为止已经给出了许多质量较好的图像。在下边的讨论中将列出几个低通滤波的实际应用。第一个例子来自机器感知领域，即字符识别应用；第二个例子来自印刷和出版业；第三个例子涉及处理卫星和航空图像。

图 4.19 显示了一幅低分辨率文本。一些打印文本就类似这样，例如，传真、复印材料、历史记录等。正如劣质文本那样，做像污点、折痕和断裂纸面那样的特殊取样并不困难。图 4.19(a) 中的放大部分显示，在这一特殊文件中，字符由于分辨率不足有了失真形状，并且许多字符断裂了。尽管人为填充这些裂缝在视觉上并不困难，但机器识别系统阅读这些断裂字符却很困难。处理这类问题常用的方法是通过模糊输入图像桥接这些裂缝。图 4.19(b) 显示了怎样使用 $D_0 = 80$ 的高斯低通滤波器的简单处理很好地修复字符。图像尺寸为 444 × 508 像素。



(a)　(b)

图 4.19　(a)低分辨率的文本样本(注意放大图中的断裂字符)，(b)用 GLPF 滤波的结果(断开的字符线段被连上了)

低通滤波是印刷和出版业的主题，在这里作为预处理函数应用，包括如 3.7.2 节讨论过的去尖锐模板。"美容"处理是在印刷中低通滤波的另一个应用。如图 4.20 所示，应用低通滤波从一幅尖锐的原像产生了平滑、看上去很柔和的外观。对于人脸，主要目的是减少皮肤细纹的锐化程度和小斑点。图 4.20(b) 和 (c) 中的放大部分明显地显示出在这种特殊情况下有效地减少了眼睛周围的细纹。事实上，平滑后的图像看上去十分柔和、美观。

图 4.21 显示了相同图像两种低通滤波的应用，但是具有不同目的。图 4.21(a) 是一幅588 × 600 的高分辨率辐射计(VHRR)图像，显示了墨西哥湾(暗的)和佛罗里达(亮的)的一部分，该图像取自 NOAA 卫星(注意水平传感器扫描线)。水体间的边界是由环形气流引起的。这幅图像是直观的遥感图像，传感器(对于多数传感器已超出了当前的讨论范畴)具有沿着一个方向产生突出扫描线的倾向，在该线上场景将被扫描。低通滤波器是减少这些线的粗略但简单的方法，如图 4.21(b) 所示(在第 5 章将考虑更有效的方法)。该图像由 $D_0 = 30$ 的高斯低通滤波器得到。这里的目的是尽可能模糊更多的细节，而保留大的可识别特征。例如，这种类型的滤波器对于图像分析系统是预处理的一部分，该系统寻找图像的边沿特征。具有这些特征的一个例子可能是给定尺寸的湖[例如佛罗里达下面东部的奥基乔比湖，如图 4.21(c) 所示的一个近似环形的暗区]。低通滤波通过消除比感兴趣特征小的特征来简化分析。



图 4.20 (a)原图像(1 028 × 732 像素),(b)用 $D_0 = 100$ 的 GLPF 滤波的结果,(c)用 $D_0 = 80$ 的 GLPF 滤波的结果,注意在(b)和(c)的放大部分细纹的减少

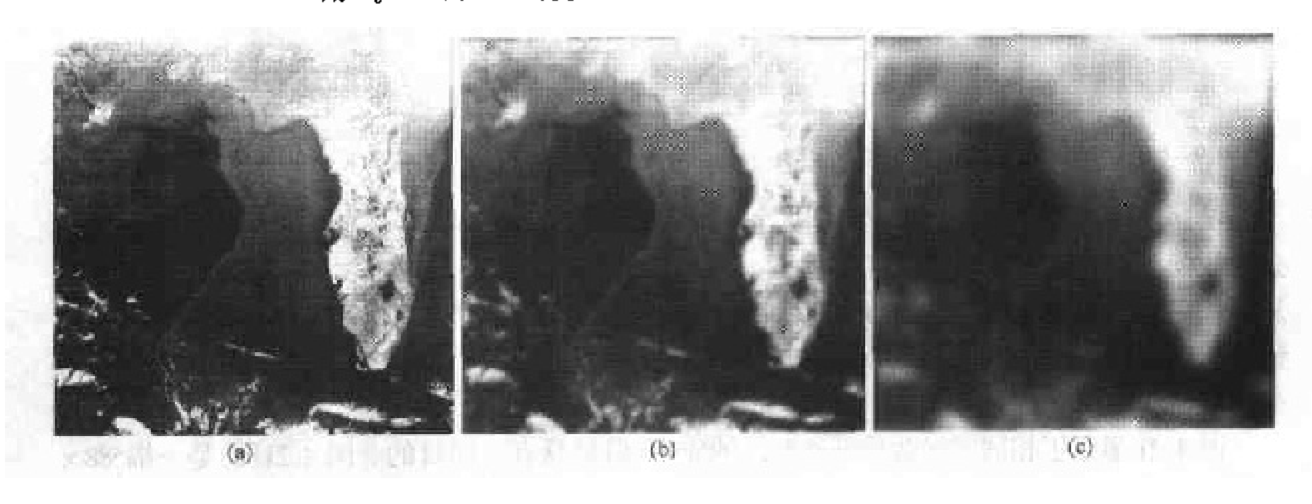


图 4.21 (a)显示突出扫描线的图像,(b)用 $D_0 = 30$ 的 GLPF 滤波的结果,(c)用 $D_0 = 10$ 的 GLPF 滤波的结果(原图像由 NASA 提供)

4.4 频率域锐化滤波器

前边的章节已经显示了通过衰减傅里叶变换的高频成分可以使图像模糊。由于在灰度级的边缘和其他地方的急剧变化与高频成分有关,图像的锐化能够在频率域用高通滤波处理实现,而衰减低频成分并不会扰乱傅里叶变换的高频信息。正如 4.3 节那样,我们认为只有零相移滤波器是放射状对称的。本节的全部滤波理论都建立在 4.2.3 小节列出过程的基础上。

本节中的滤波器函数是对前面章节讨论的理想低通滤波器的精确反操作，即本节讨论的高通滤波器的传递函数可由下面的关系式得到：

$$H_{hp}(u,v) = 1 - H_{lp}(u,v) \tag{4.4.1}$$

这里，$H_{lp}(u,v)$是相应低通滤波器的传递函数。也就是说，被低通滤波器衰减的频率能通过高通滤波器，反之亦然。

在本节中，我们考虑的是理想的、巴特沃思型和高斯型三种高通滤波器。正如前面章节所示，在频率域和时域分别说明这些滤波器的特性。图 4.22 显示了这些滤波器典型的三维剖面图、图像表示和横截面。像前边一样，我们看到巴特沃思型滤波器为理想滤波器的尖锐化和高斯型滤波器的完全光滑之间的一种过渡。图 4.23 说明了这些滤波器在空间域的形状。回忆一下，一个频率域滤波器的空间域表达式是这样得来的：(1)用$(-1)^{u+v}$乘以 $H(u,v)$中心化；(2)计算反 DFT；然后(3)将反 DFT 变换的实部乘以$(-1)^{x+y}$。这些图的重要性质将在本节后面部分讨论。

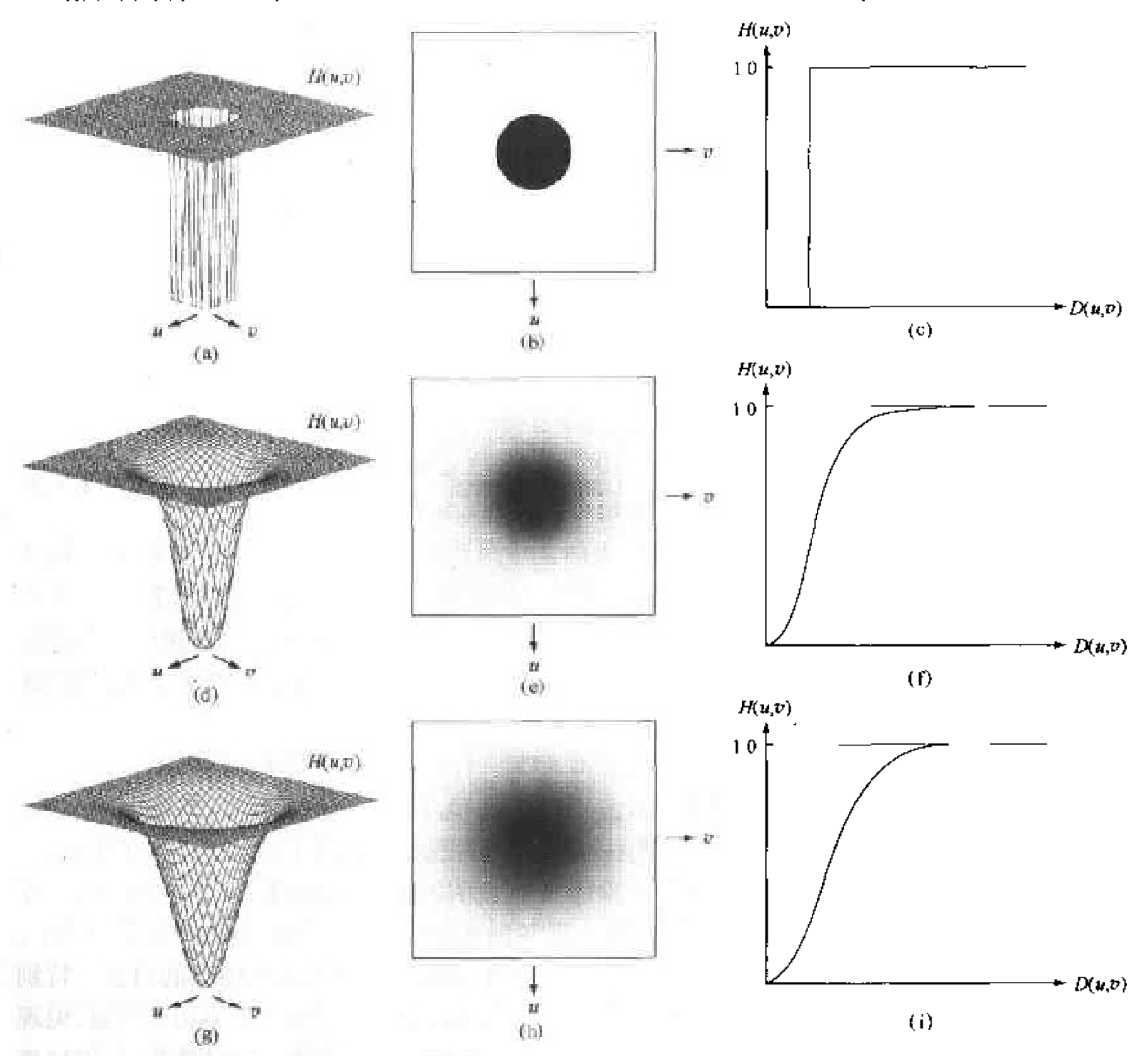


图 4.22　最上面一行：透视图、图像表示和典型的理想高通滤波器的横截面，中间和底下一行：典型的巴特沃思和高斯高通滤波器的透视图、图像表示和横截面

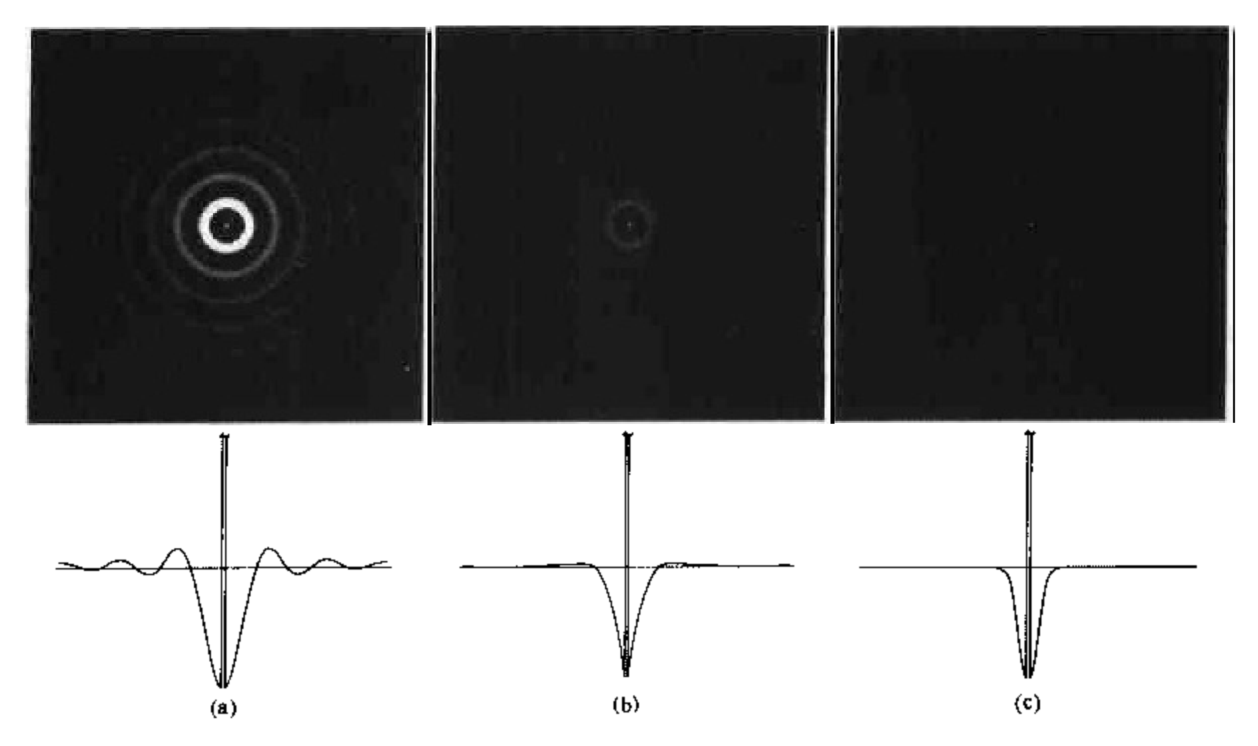


图 4.23 典型的空间表示。(a)理想的,(b)巴特沃思型和(c)高斯频域高通滤波器及相应的灰度剖面图

4.4.1 理想高通滤波器

一个二维理想高通滤波器(IHPF)的定义如下:

$$H(u,v)=\begin{cases}0 & D(u,v)\leqslant D_0\\ 1 & D(u,v)>D_0\end{cases} \tag{4.4.2}$$

D_0 是从频率矩形的原点测得的截止长度,$D(u,v)$ 由式(4.3.3)给出。这个表达式直接从式(4.4.1)和式(4.4.2)得到。如预料的那样,这个滤波器是与低通滤波器相对的,它将以 D_0 为半径的圆周内的所有频率置零,而毫不衰减地通过圆周外的任何频率。如同理想低通滤波器的情况,IHPF 也是物理上用电子元件无法实现的。但是,因为它可以用计算机来实现,所以认为它是完备的,这里的讨论也很简要。

因为 IHPF 与式(4.4.1)有关,我们预料它和 ILPF 一样有相同的振铃性质[见图 4.23(a)]。这由图 4.24 可以清楚地看出来,图中把 D_0 分别设置为 15,30 和 80 个像素点,使用图 4.11(a)的原始图像,产生了不同的 IHPF 结果。图 4.24(a)的振铃现象太严重了以至于产生了失真,物体的边界也被加粗了(如那个大字母"a")。顶部 3 个圆的边缘不是很清晰(这 3 个物体的灰度很接近于背景灰度,产生微小的幅度不连续)。观察图 4.23(a) IHPF 空间表示的"点"的大小,空间域的滤波是滤波器与图像的卷积,因此微小的物体和线条显现出几乎纯粹的白色。特别注意顶行的 3 个小方形和细直条。当 $D_0=30$ 时情况有所改善。虽然边缘失真仍旧明显,但现在开始看到对于微小物体的过滤。由于现在很熟悉的频率域和空间域的相对关系,我们知道这种滤波器的点的大小比 $D_0=5$ 的滤波器更小。在 $D_0=80$ 时,结果更像一个高通滤波过的图像。在这里,边缘更加清晰,失真更小,而且细小的物体也能得到正确的过滤。

4.4.2　巴特沃思高通滤波器

n 阶且截止频率距原点的距离为 D_0 的巴特沃思型高通滤波器(BHPF)的传递函数由下式给出：

$$H(u,v)=\frac{1}{1+[D_0/D(u,v)]^{2n}} \tag{4.4.3}$$

其中 $D(u,v)$ 由式(4.3.3)给出。式(4.4.3)由式(4.1.1)和式(4.3.6)直接得到。图 4.22 中间一行显示了 BHPF 函数的图像和横截面。

同低通滤波器的情况一样，可以认为巴特沃思型高通滤波器比 IHPF 更平滑。二阶且 D_0 值同图 4.24 的设置相同的 BHPF 特性示于图 4.25。边缘失真比图 4.24 小得多，甚至对于最小的截止频率也一样。一旦 IHPF 和 BHPF 中点的大小相同[见图 4.23(a)和(b)]，这两个滤波器过滤小目标的性能便能够比较了。若转变为高截频值 BHPF 更平滑。

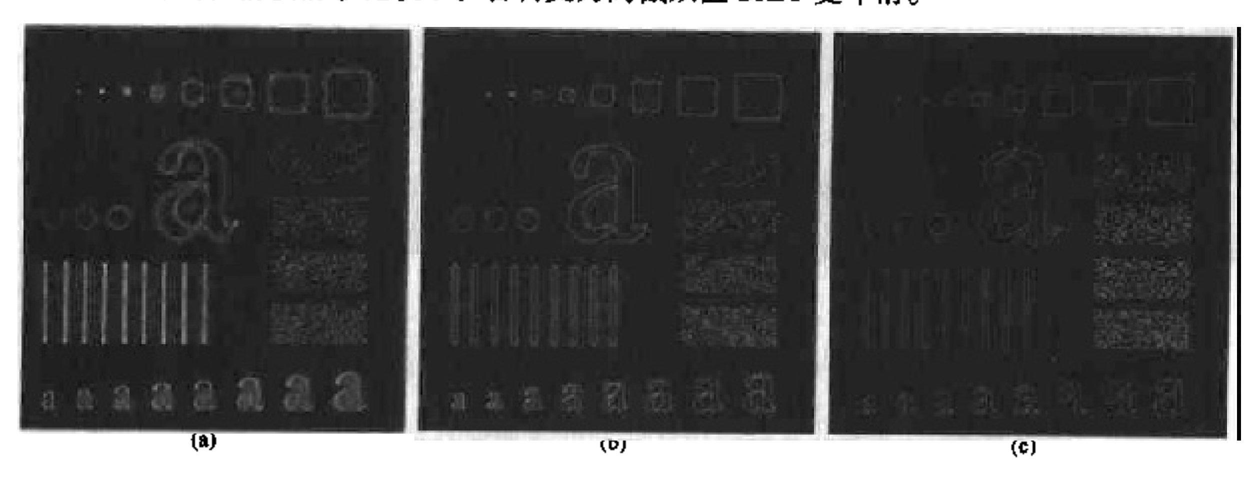


图 4.24　理想高通过滤图 4.11(a)中图像的结果，$D_0=15$，30和80。振铃问题在(a)和(b)中十分明显

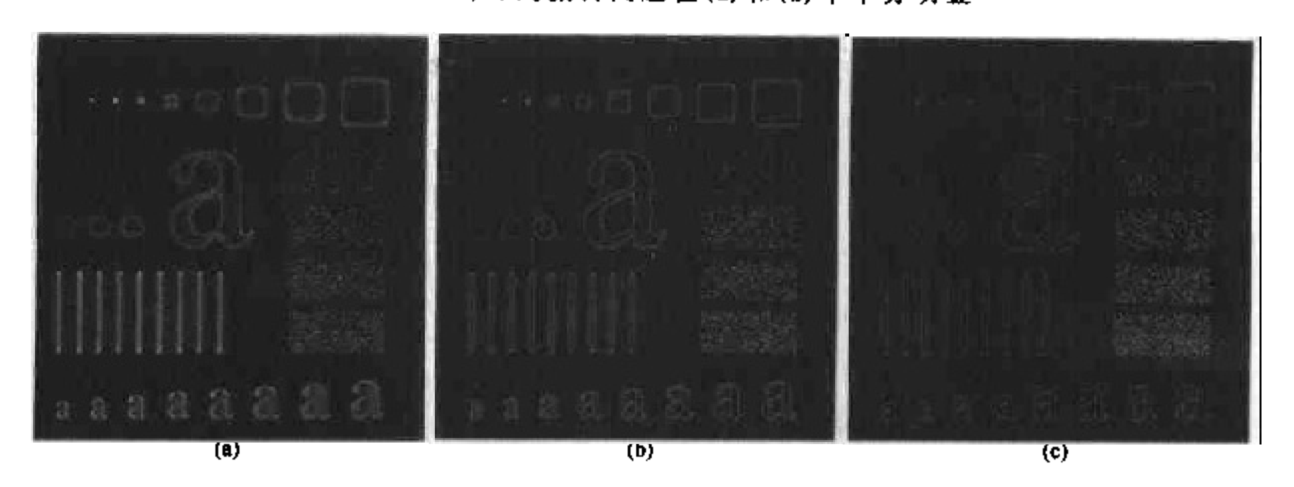


图 4.25　用二阶且 $D_0=15,30$ 和 80 的 BHPF 对图 4.11(a)中图像进行高通滤波的结果。其结果比用ILPF得到的结果平滑得多

4.4.3 高斯型高通滤波器

截频距原点为 D_0的高斯型高通滤波器(GHPF)的传递函数是:

$$H(u,v) = 1 - e^{-D^2(u,v)/2D_0^2} \tag{4.4.4}$$

其中 $D(u,v)$由式(4.3.3)给出。该公式可由式(4.1.1)和式(4.3.8)直接得到。图 4.22 第三行显示了 GHPF 函数的透视图、图像和横截面。沿用 BHPF 相同的形式,在图 4.26 中显示了用 GHPF 的可比较的结果。正如所料,得到的结果比前两个滤波器更平滑。即使是对微小物体和细条用高斯滤波器过滤也是较清晰的。

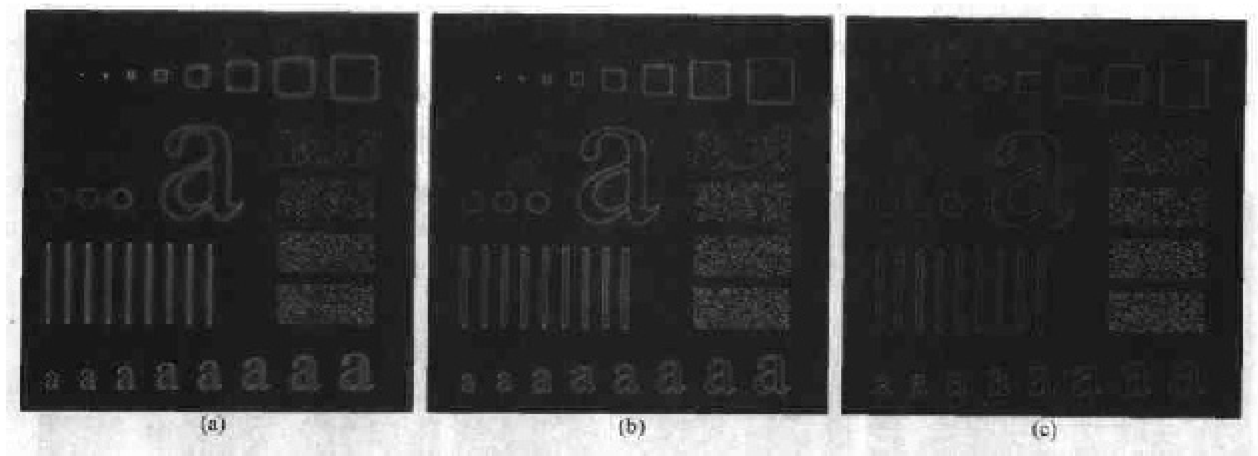


图 4.26 用一个二阶、D_0 分别等于 15,30 和 80 的 GHPF 对图像进行高通滤波的结果。可与图 4.24 和图 4.25 相比较

如 4.2.4 小节所述,高通滤波器能够用高斯型低通滤波器的差分构成。这些不同的滤波器有着更多的参数,因此能够对滤波器的形状进行更多控制。但是,式(4.4.4)给出的简单滤波器在实践中通常已足够用,而且从经验上来说更容易公式化。

4.4.4 频率域的拉普拉斯算子

公式表示如下:

$$\Im\left[\frac{d^n f(x)}{dx^n}\right] = (ju)^n F(u) \tag{4.4.5}$$

从这个简单的表达式可以得到:

$$\Im\left[\frac{\partial^2 f(x,y)}{\partial x^2} + \frac{\partial^2 f(x,y)}{\partial y^2}\right] = (ju)^2 F(u,v) + (jv)^2 F(u,v) \tag{4.4.6}$$

$$= -(u^2+v^2)F(u,v)$$

式(4.4.6)右边括弧内的表达式就是在式(3.7.1)中定义的拉普拉斯算子。这样,我们就得到 $f(x,y)$一个重要的结论:

$$\Im[\nabla^2 f(x,y)] = -(u^2+v^2)F(u,v) \tag{4.4.7}$$

即,频率域的拉普拉斯算子可以由如下滤波器实现:

$$H(u,v) = -(u^2+v^2) \tag{4.4.8}$$

正如本章的所有滤波运算那样,前提都是 $F(u,v)$的原点在进行图像变换前已由执行运算

$f(x,y)(-1)^{x+y}$中心化了。正如前面讨论过的,如果 f(和 F)是 $M \times N$ 大小,这种运算移动了变换中心,则$(u,v)=(0,0)$就是在频率矩形中的点$(M/2,N/2)$。如前所述,滤波函数中心也需要被平移为:

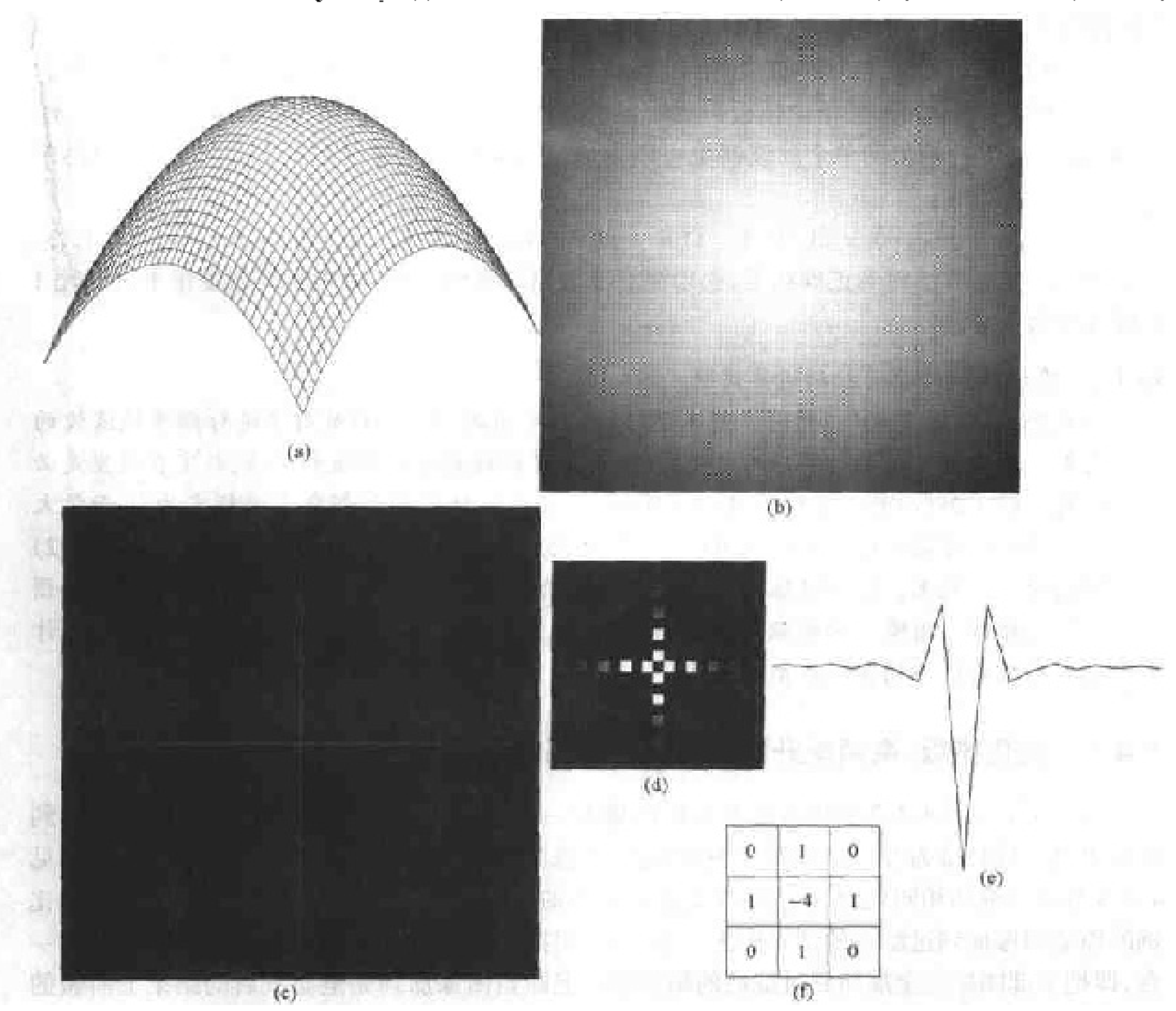
$$H(u,v) = -[(u-M/2)^2+(v-N/2)^2] \tag{4.4.9}$$

空间域拉普拉斯算子过滤后的图像可由计算 $H(u,v)F(u,v)$的反傅里叶变换得到:

$$\nabla^2 f(x,y) = \Im^{-1}\{-[(u-M/2)^2+(v-N/2)^2]F(u,v)\} \tag{4.4.10}$$

相对地,用式(3.7.1)计算空间域拉普拉斯算子,而计算结果的傅里叶变换等价于 $F(u,v)$与 $H(u,v)$相乘。我们以熟悉的傅里叶变换对表示这种双向关系:

$$\nabla^2 f(x,y) \Leftrightarrow -[(u-M/2)^2+(v-N/2)^2]F(u,v) \tag{4.4.11}$$



0	1	0
1	-4	1
0	1	0

图 4.27　(a)频率域拉普拉斯的三维图,(b)(a)的图像表示,(c)从(b)的傅里叶反变换得到的空间域拉普拉斯,(d)(c)中原图像的放大图像,(e)通过(d)中心的灰度剖面图,(f)在 3.7 节使用的拉普拉斯模板

空间域拉普拉斯滤波函数由式(4.4.9)的反傅里叶变换得到,而且如图 4.27 所示,有一些有趣的特性。图 4.27(a)是式(4.4.9)的三维透视图。函数的中心在$(M/2,N/2)$,并且在圆顶

处的值为零,其他所有值为负值。图 4.27(b)以图像形式显示了 $H(u,v)$,而且已中心化。图 4.27(c)是空间域的拉普拉斯算子,由 $H(u,v)$乘以$(-1)^{u+v}$,再进行傅里叶反变换,然后结果的实部乘上$(-1)^{x+y}$得到。图 4.27(d)是图 4.27(c)原图的放大图。图 4.27(e)是通过放大区中心的水平灰度剖面图。最后图 4.27(f)所示的是在 3.7 节用到过的模板,它用来实现式(3.7.4)定义的空间域离散拉普拉斯算子。通过模板中心的水平剖面线与图 4.27(e)有着相同的基本形状(即两个小的正值之间是一个负值)。值得注意的是,我们是从频率域开始推导出和图 3.39(a)有着相似形式的空间域模板。

在式(3.7.5)中,从原始图形中减去拉普拉斯算子部分,形成 $g(x,y)$的增强图像:

$$g(x,y) = f(x,y) - \nabla^2 f(x,y) \tag{4.4.12}$$

将拉普拉斯算子从原图中减去(而不是加)是由式(4.4.8)的负号决定的。这与空间域给出的式(3.7.5)相一致。可以得到相同的结论,即,图 4.27(e)的中心峰值为负值,而它的邻近点为正值。

如同在空间域那样,用单个模板得到增强的图像,在频率域中由 $H(u,v)=[1-[(u-M/2)^2+(v-N/2)^2]]$给出的单个滤波器也可以完成全部操作。在这种情况下,增强图像可用单个反变换运算得到:

$$g(x,y) = \Im^{-1}\{[1-((u-M/2)^2+(v-N/2)^2)]F(u,v)\} \tag{4.4.13}$$

为了用这一实现方法得到正确结果,必须注意滤波器的缩放比例,因为滤波器变量平方后比 1 大好几个数量级。

例 4.7 在频率域的拉普拉斯运算说明

图 4.28(a)与图 3.40(a)相同。图 4.28(b)显示了用式(4.4.10)对图像进行频率域滤波的结果。作为拉普拉斯过滤后的典型图像,具有可比较数量的正值和负值,标定在这里是必需的。图 4.28(c)显示了标定过的图像(为了显示),故它的大部分负值标定为 0,而最大正值标定为可显示的最大灰度值(此例中为 255)。最后,图 4.28(d)显示了用式(4.4.12)得到的增强效果。微小特征细节的锐化增强很明显,正如同用拉普拉斯算子所得到的预期效果那样。图像显示的顺序与图 3.40 相比完全相同,只不过是采用空间域的技术计算。结果对所有的实际应用都相同。

4.4.5 钝化模板、高频提升滤波和高频加强滤波

从 4.4.1 节到 4.4.3 节所有滤波后的图像有一个共同特点:它们背景的平均强度减小到接近黑色。这归结为在这些图像上所用的高通滤波器除去了傅里叶变换的零频率成分(见 4.2.3 节所讨论的相同现象)。同 3.7.2 节所讨论的一样,对这个问题的解决办法是把一定比例的原始图像加到过滤后的结果中去。事实上,用拉普拉斯算子增强确实能准确地做到这一点,即把全部图像完全加回到过滤后的结果中。把原始图像加到完全滤波后的结果上所做的贡献很有益。这种处理叫做高频提升过滤,它是钝化模板的推广。在 3.7.2 节中介绍过这些概念。在这里用频率域的概念和表示法重复一遍。

钝化模板简单地由从一幅图像减去其自身模糊图像而生成的锐化图像构成。采用频域技术,这意味着从图像自身减去低通滤波后的图像而得到高通滤波的图像。即:

$$f_{hp}(x,y) = f(x,y) - f_{lp}(x,y) \tag{4.4.14}$$

高频提升过滤通过将 $f(x,y)$乘上一个常数 $A \geq 1$ 而产生:

$$f_{hb} = Af(x,y) - f_{lp}(x,y) \tag{4.4.15}$$

这样,高频提升过滤为我们提升整幅图像的增强效果提供了一定灵活性。这一公式可写为下式:

$$f_{hb}(x,y) = (A-1)f(x,y) + f(x,y) - f_{lp}(x,y) \tag{4.4.16}$$

然后,由式(4.4.14)可得:

$$f_{hb}(x,y) = (A-1)f(x,y) + f_{hp}(x,y) \tag{4.4.17}$$

这个结果基于一幅高通图像而不是低通图像。当 $A=1$ 时,高频提升过滤变为常规的高通过滤。当 A 增加到超过 1 时,图像自身的贡献显得更加显著。

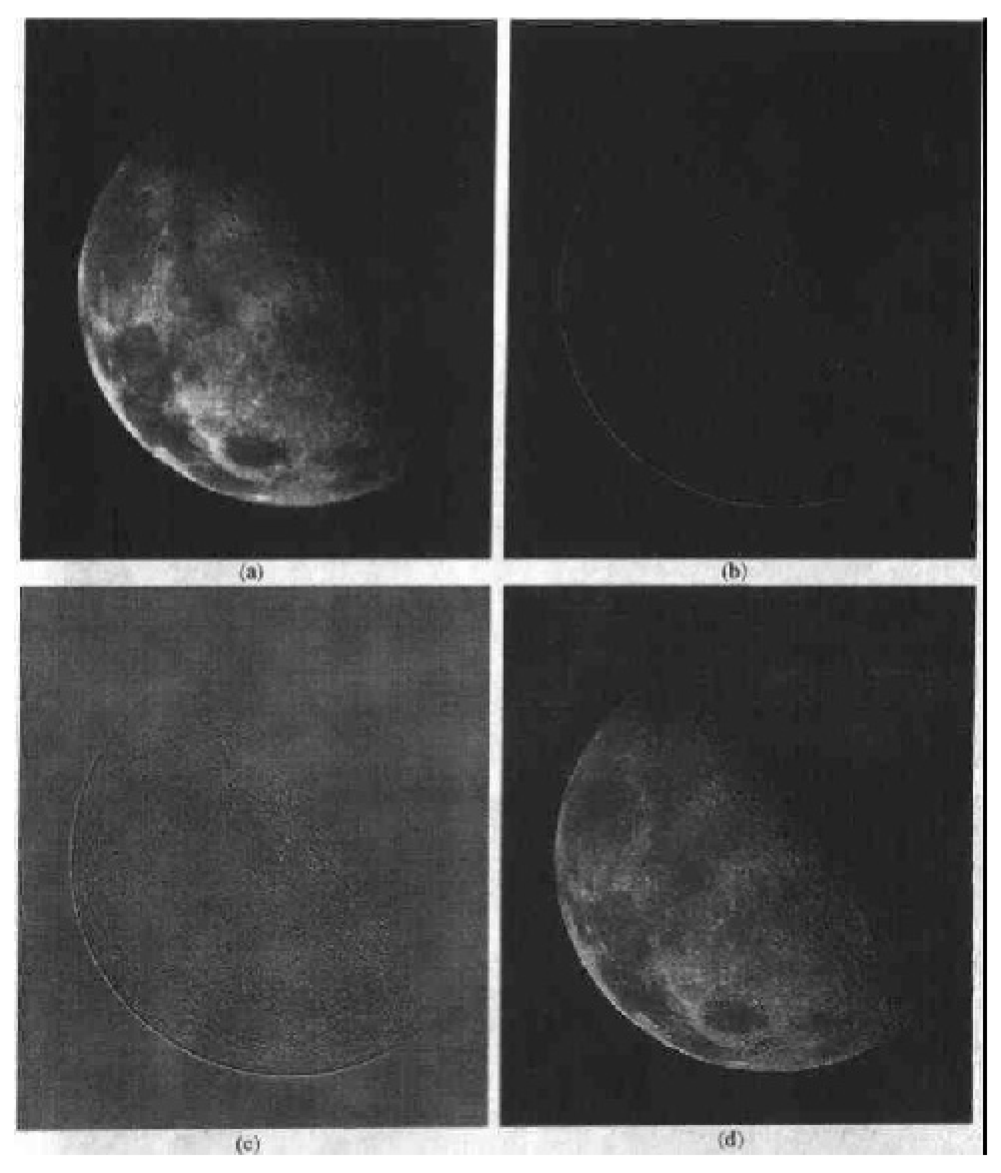


图 4.28　(a)月球北极图像,(b)拉普拉斯滤波后的图像,(c)标定后的图像,(d)用式(4.4.12)增强的图像(原图像由 NASA 提供)

从式(4.4.14)可知,$F_{hp}(u,v) = F(u,v) - F_{lp}(u,v)$,而 $F_{lp}(u,v) = H_{lp}(u,v)F(u,v)$,这里 H_{lp}是低通滤波器的传递函数。因此,钝化模板在频率域中可由混合滤波器直接执行:

$$H_{hp}(u,v) = 1 - H_{lp}(u,v) \tag{4.4.18}$$

注意这个结果同式(4.4.1)相一致。类似地,高频提升过滤也可以由混合型滤波器执行:

$$H_{hb}(u,v) = (A-1) + H_{hp}(u,v) \tag{4.4.19}$$

这里 $A \geqslant 1$。这个过程由滤波器乘上输入图像的变换(中心化后),再将得到的结果进行反变换组成。该结果的实部乘上$(-1)^{x+y}$得到空间域高频提升滤波后的图像$f_{hb}(x,y)$。

例 4.8 频率域的高频提升滤波

图 4.29 显示了与图 3.43 相同的图像序列,但是使用频域计算。图 4.29(a)是输入图像,图 4.29(b)是高频提升滤波后的图像。在这个例子中为了与图 3.43 那些图像比较,使用拉普拉斯作为高通滤波器,并用式(4.4.10)计算。不选择混合滤波器做拉普拉斯的标定更容易。图4.29(c)的图像是 $A=2$ 时用式(4.4.17)得到的。同图 3.43(c)一样,这幅图像更加锐利,但仍旧太暗。图 4.29(d)是 $A=2.7$ 时得到的,在效果上意味着拉普拉斯算子被减去前,输入图像乘以 1.7。与图 3.43 一样,这是一个改进后的结果。然而,图 4.29(d)不如图 4.43(d)那样锐利。原因是拉普拉斯算子的频率域表达式更接近于排除了对角线邻域的模板[见图 4.27(f)]。从图 3.41 的例子可知包含对角线邻域的模板产生微小的更加锐化的结果。这种差异当细节很小时(如同月球的例子)通常无法注意到,但是,对有较多特性的图像,差异将变得很显著。

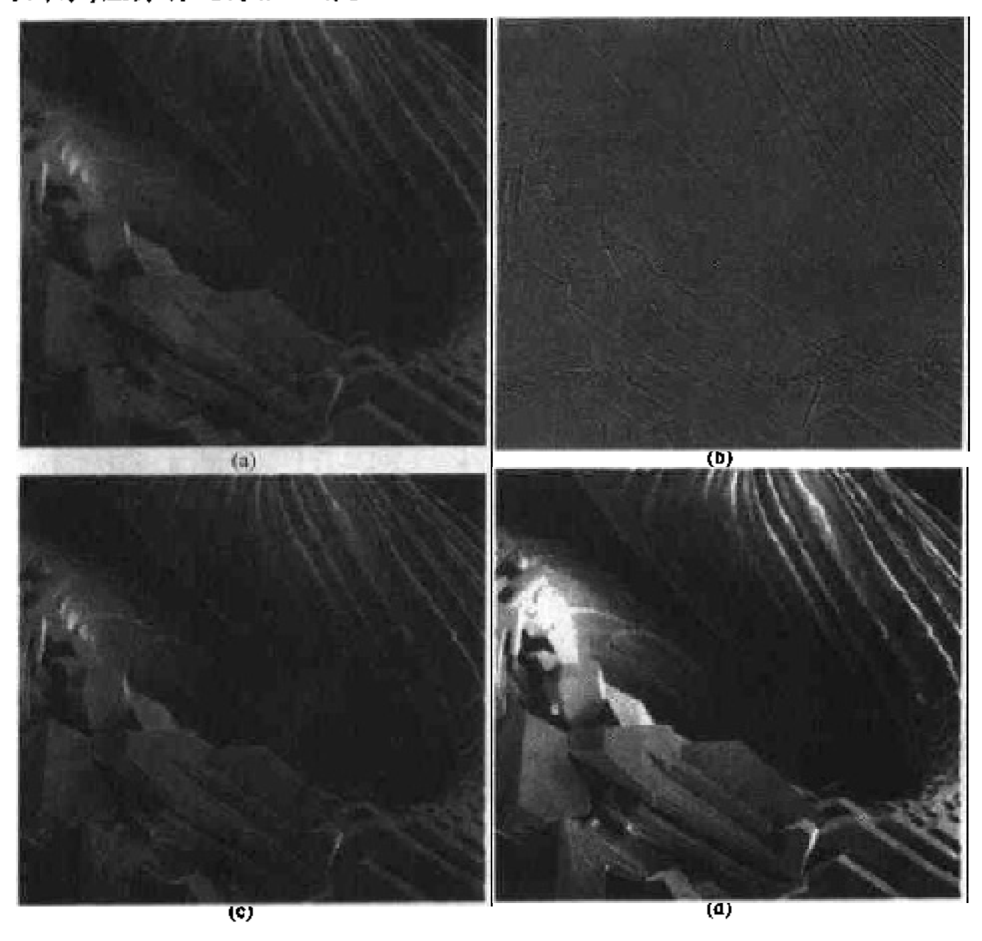


图 4.29 与图 3.43 相同,但用频率域滤波。(a)输入图像,(b)(a)的拉普拉斯图像,(c)$A=2$时用式(4.4.17)得到的图像,(d)与(c)相同,但$A=2.7$(原图像由俄勒冈大学地质系的Michael Shaffer先生提供)

有时,用一幅图像的高频成分强调增强的作用是有益的。这种情况下,在高通滤波器函数

前简单地乘以一个常数，再增加一个偏移以便使零频率不被滤波器除掉。这种处理叫做高频加强，有如下传递函数：

$$H_{hfe}(u,v) = a + bH_{hp}(u,v) \tag{4.4.20}$$

这里，$a \geqslant 0$ 且 $b > a$。a 的典型值在 0.25 到 0.5 之间，b 的典型值在 1.5 到 2.0 之间。参考式(4.4.17)，可知，当 $a = (A-1)$ 且 $b = 1$ 时高频加强转化为高频提升过滤。当 $b > 1$ 时，高频得到加强，这就是它名字的由来。

例 4.9　高频加强滤波

图 4.30(a)显示了一个窄灰度级的 X 射线胸透图。这里主要的目的是锐化图像。X 射线不能用透镜聚焦的方式聚焦，而且产生的图像通常容易有轻微的模糊。虽然这幅特殊图像的灰度偏向灰度级较暗的一端，在这里也能给出用空间域处理来实现频率域过滤的方法的例子。

图 4.30(b)显示了用二阶、D_0 等于垂直方向值的 5% 的巴特沃思型滤波器进行高通滤波的结果。只要滤波器的半径不太小，频率接近变换后的原点，高通滤波就不会对参数过度敏感。如预料的那样，过滤后的结果毫无特征，但它显示出微弱的主要图像边缘。高频加强过滤(此例中 $a = 0.5$ 且 $b = 2.0$)的优势显示于图 4.30(c)中。虽然图像仍然很暗，灰度色调由于低频分量的保持没有丢失。

如 3.3 节所述，以灰度为特征的图像在很窄的灰度范围内对于直方图均衡是理想选择。正如图 4.30(d)所示的那样，这确实是进一步增强图像的一个合适的方法。请注意，骨骼结构的清晰度和在另外三幅图像中都看不见的其他细节。最后增强的图像存在一些噪声，但这是在灰度范围扩大时射线图像的典型现象。用高频加强和直方图均衡混合的方法得到的结果比单独采用其中任一种方法产生的效果要好。

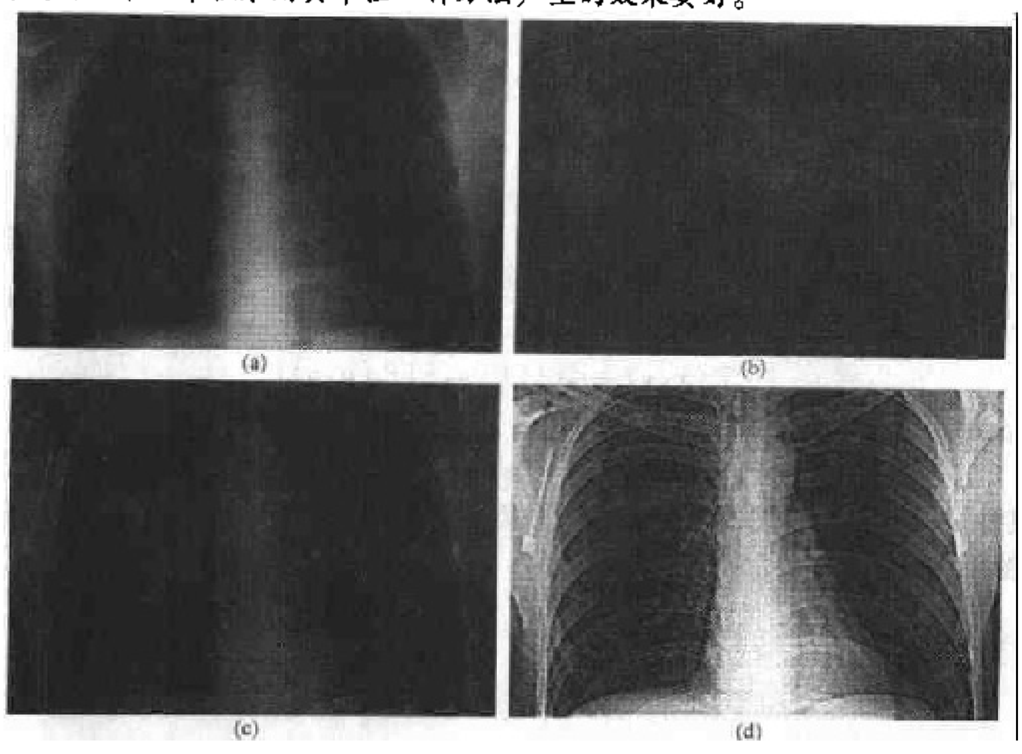


图 4.30　(a)胸部 X 光图像，(b)巴特沃思高通滤波的结果，(c)高频增强滤波的结果，(d)对(c)执行直方图均衡的结果（原图像由密歇根大学医学院解剖科学部的 Thomas R. Gest 博士提供）

4.5 同态滤波器

2.3.4节介绍的照度-反射模型可用来开发一种频域处理,通过同时进行灰度范围的压缩和对比度增强来改进一幅图像的外观。从2.3.4节的讨论中可知图像$f(x,y)$能被表达成照度和反射两部分的乘积:

$$f(x,y) = i(x,y)r(x,y) \tag{4.5.1}$$

式(4.5.1)不能用来直接对照度和反射的频率部分分别进行操作,原因是两个函数乘积的傅里叶变换是不可分的。换而言之:

$$\Im\{f(x,y)\} \neq \Im\{i(x,y)\}\Im\{r(x,y)\}$$

然而,假设定义:

$$\begin{aligned} z(x,y) &= \ln f(x,y) \\ &= \ln i(x,y) + \ln r(x,y) \end{aligned} \tag{4.5.2}$$

那么

$$\begin{aligned} \Im\{z(x,y)\} &= \Im\{\ln f(x,y)\} \\ &= \Im\{\ln i(x,y)\} + \Im\{\ln r(x,y)\} \end{aligned} \tag{4.5.3}$$

或

$$Z(u,v) = F_i(u,v) + F_r(u,v) \tag{4.5.4}$$

这里,$F_i(u,v)$和$F_r(u,v)$分别是$\ln i(x,y)$和$\ln r(x,y)$的傅里叶变换。

如果借助于一个滤波函数$H(u,v)$处理$Z(u,v)$,那么从式(4.2.27)可得:

$$\begin{aligned} S(u,v) &= H(u,v)Z(u,v) \\ &= H(u,v)F_i(u,v) + H(u,v)F_r(u,v) \end{aligned} \tag{4.5.5}$$

这里,$S(u,v)$是结果的傅里叶变换。在空间域,

$$\begin{aligned} s(x,y) &= \Im^{-1}\{S(u,v)\} \\ &= \Im^{-1}\{H(u,v)F_i(u,v)\} + \Im^{-1}\{H(u,v)F_r(u,v)\} \end{aligned} \tag{4.5.6}$$

令

$$i'(x,y) = \Im^{-1}\{H(u,v)F_i(u,v)\} \tag{4.5.7}$$

和

$$r'(x,y) = \Im^{-1}\{H(u,v)F_r(u,v)\} \tag{4.5.8}$$

则式(4.5.6)可以用如下形式表示:

$$s(x,y) = i'(x,y) + r'(x,y) \tag{4.5.9}$$

最后,只要$z(x,y)$是原始图像$f(x,y)$取对数得到的,相反的(取指数)操作就能产生符合要求的增强图像,由$g(x,y)$表示,也就是:

$$\begin{aligned} g(x,y) &= e^{s(x,y)} \\ &= e^{i'(x,y)} \cdot e^{r'(x,y)} \\ &= i_0(x,y)r_0(x,y) \end{aligned} \tag{4.5.10}$$

这里

$$i_0(x,y) = e^{i'(x,y)} \tag{4.5.11}$$

$$r_0(x,y) = e^{r'(x,y)} \tag{4.5.12}$$

是输出图像的照射分量和反射分量。

采用上述概念的增强方法总结于图 4.31 中。这个方法是基于称为同态系统的一类系统的特例。在这个特殊应用中，方法的关键是用式(4.5.4)的形式将照射分量和反射分量分开。同态滤波函数 $H(u,v)$能分别对两个分量进行操作，如式(4.5.5)指出的那样。

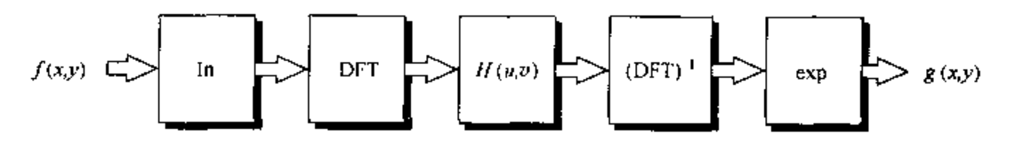


图 4.31　图像增强中的同态滤波

图像照射分量通常以空间域的慢变化为特征，而反射分量往往引起突变，特别在不同物体的连接部分。这些特性导致图像对数的傅里叶变换的低频成分与照度相联系，而高频成分与反射联系在一起。虽然这些联系只是大体上的近似，但它们用于图像增强时是有益的。

许多控制能通过用同态滤波器对照射分量和反射分量操作来加强。这些控制需要一个滤波器函数 $H(u,v)$来规范。它能以不同的方法影响傅里叶变换的高低频成分。图 4.32 显示了这种滤波器的横截面图。一旦 γ_L 和 γ_H 选定，而 $\gamma_L < 1$ 且 $\gamma_H > 1$，图 4.32 所示的滤波器函数往往减少低频(照度)的贡献，而增加高频(反射)的贡献。结果是同时进行动态范围的压缩和对比度的增强。

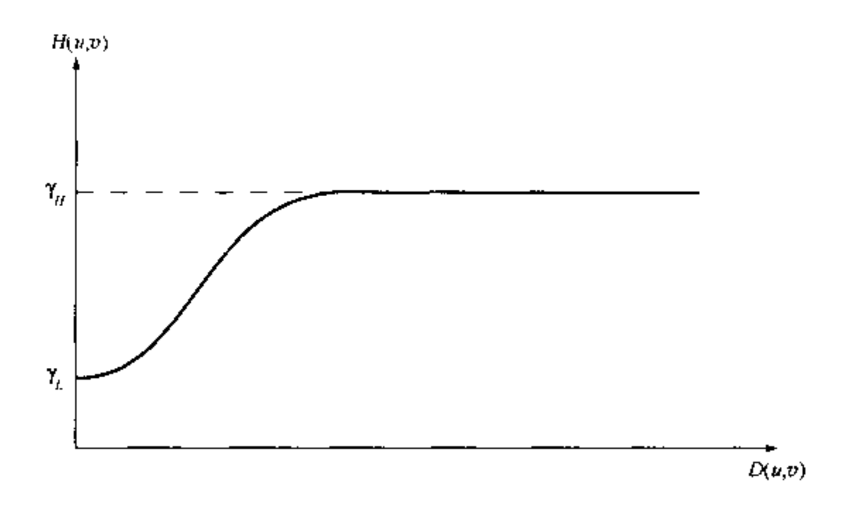


图 4.32　一个圆对称滤波函数的横截面。$D(u,v)$是中心变换后距原点的距离

图 4.32 所示的曲线形状能用前面章节所述的任何一种理想高通滤波器的基本形式近似。例如，采用高斯型高通滤波器稍微修改过的形式：

$$H(u,v) = (\gamma_H - \gamma_L)[1 - e^{-c(D^2(u,v)/D_0^2)}] + \gamma_L \tag{4.5.13}$$

这里，$D^2(u,v)$由式(4.3.3)给出，常数 c 被引入用来控制滤波器函数斜面的锐化，它在 γ_L 和 γ_H 之间过渡。这类滤波器与 4.4 节末尾讨论的高频加强滤波器相似。

例 4.10　用同态滤波增强

图 4.33 是用图 4.32 所示的同态滤波器函数得到的典型结果。图 4.33(a)所示的原始图

像中,掩体内部的细节被墙外的光所遮蔽。图 4.33(b)显示了用同态滤波器处理这一图像的结果,其中图 4.32 的滤波函数中的 $\gamma_L = 0.5$ 且 $\gamma_H = 2.0$。减小亮度的动态范围,增强对比度可以使掩体内的细节显示出来,并且平衡墙外的灰度。增强后的图像同样更加锐化。

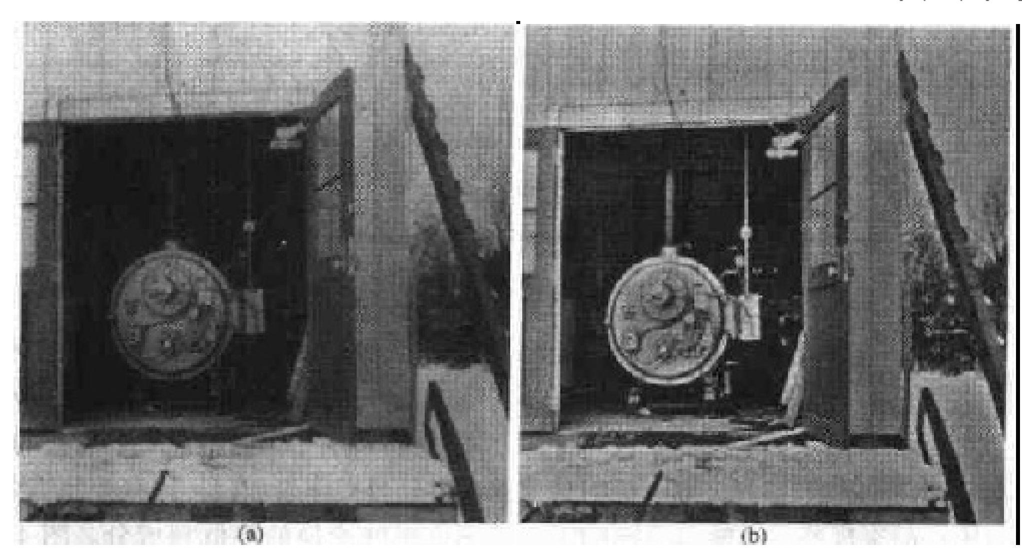


图 4.33 (a)原图像,(b)用同态滤波处理后的图像(注意掩体内的细节)(Stockham)

4.6 实现

这一节讨论傅里叶变换的基础知识。从对一些二维傅里叶变换特性的总结开始,通过快速傅里叶变换(FFT)的简要复习推出结论。

4.6.1 一些二维傅里叶变换的性质

平移

傅里叶变换对有如下平移性质:

$$f(x,y)e^{j2\pi(u_0 x/M + v_0 y/N)} \Leftrightarrow F(u-u_0, v-v_0) \tag{4.6.1}$$

$$f(x-x_0, y-y_0) \Leftrightarrow F(u,v)e^{-j2\pi(ux_0/M + vy_0/N)} \tag{4.6.2}$$

这里,如式(4.2.31)中介绍的那样,引入双箭头用于表示傅里叶变换对。当 $u_0 = M/2$ 且 $v_0 = N/2$ 时有:

$$e^{j2\pi(u_0 x/M + v_0 y/N)} = e^{j\pi(x+y)} = (-1)^{x+y}$$

在这种情况下,式(4.6.1)变为:

$$f(x,y)(-1)^{x+y} \Leftrightarrow F(u-M/2, v-N/2) \tag{4.6.3}$$

类似地,

$$f(x-M/2, y-N/2) \Leftrightarrow F(u,v)(-1)^{(u+v)} \tag{4.6.4}$$

可见式(4.6.3)与式(4.2.21)相同,式(4.2.21)用于中心化变换。这些结果基于变量 u 和 v,它们的值分别在$[0, M-1]$和$[0, N-1]$范围内。在计算机执行的过程中,变量 u 的值从 1 变到

M,而变量 v 的值从 1 变到 N,在这种情况下,变换的实际中心变为 $u=(M/2)+1, v=(N/2)+1$。

分配性和比例变换性

从傅里叶变换的定义有:

$$\Im[f_1(x,y)+f_2(x,y)] = \Im[f_1(x,y)] + \Im[f_2(x,y)] \tag{4.6.5}$$

且通常

$$\Im[f_1(x,y)\cdot f_2(x,y)] \neq \Im[f_1(x,y)]\cdot\Im[f_2(x,y)] \tag{4.6.6}$$

换而言之,傅里叶变换对加法具有分配性,而不适用于乘法。傅里叶反变换适用于相同的结论。类似地,对于比例因子 a 和 b,可得:

$$af(x,y) \Leftrightarrow aF(u,v) \tag{4.6.7}$$

和

$$f(ax,by) \Leftrightarrow \frac{1}{|ab|}F(u/a,v/b) \tag{4.6.8}$$

旋转

如果引入极坐标:

$$x = r\cos\theta, y = r\sin\theta, u = \omega\cos\varphi, v = \omega\sin\varphi$$

那么 $f(x,y)$ 和 $F(u,v)$ 分别变成 $f(r,\theta)$ 和 $F(\omega,\varphi)$。在傅里叶变换的定义中直接替换得到:

$$f(r,\theta+\theta_0) \Leftrightarrow F(\omega,\varphi+\theta_0) \tag{4.6.9}$$

这个表达式指出以角度 θ_0 旋转 $f(x,y)$,$F(u,v)$ 将转过相同的角度。类似地,旋转 $F(u,v)$,$f(x,y)$ 将转过相同的角度。

周期性和对称性

离散傅里叶变换有如下周期性性质:

$$F(u,v) = F(u+M,v) = F(u,v+N) = f(u+M,v+N) \tag{4.6.10}$$

反变换也是周期性的:

$$f(x,y) = f(x+M,y) = f(x,y+N) = f(x+M,y+N) \tag{4.6.11}$$

共轭对称的概念在 4.2 节中介绍过,为便于学习,在这里再介绍一下:

$$F(u,v) = F^*(-u,-v) \tag{4.6.12}$$

从中可以得到结论:频谱也是关于原点对称的:

$$|F(u,v)| = |F(-u,-v)| \tag{4.6.13}$$

这些等式的有效性是建立在式(4.2.16)和式(4.2.17)上的。

周期性的重要性在图 4.34(a)中得到说明,图中显示了一维变换 $F(u)$ 的频谱[见式(4.2.5)和式(4.2.10)]。根据式(4.6.10),$F(u)=F(u+M)$,从而 $|F(u)|=|F(u+M)|$。同样,根据式(4.6.13),$|F(u)|=|F(-u)|$。周期性指出 $F(u)$ 的周期长度为 M,对称性指出频谱关于原点对称,如图 4.34(a)所示。这幅图和前面的讨论表明从 $(M/2)+1$ 到 $M-1$ 变换值的大小是对原点左边半周期值的映射。因为离散傅里叶变换是对区间 $[0,M-1]$ 中的 u 值表述的,这个变换的结果在该区间产生两个背靠背的半周期。要显示一个完全的周期,必须将变换的原点移至点 $u=M/2$,如图 4.34(b)所示。为做到这一点,在变换前简单地将 $f(x)$ 乘以

$(-1)^{x}$,如式(4.6.3)所示。

二维频谱以相同的方法加以分析。图 4.34(c)和(d)是中心化前后(只显示一个周期)的典型频谱图。此例中,变换中心在图的左上角,而四个二维周期分量从图像的四个角背靠背地扩展到中心,且包含了高频成分。相对比,用式(4.6.3)获得的中心化变换的原点(零频率成分)在图 4.34(d)的中心,而它的各部分以共轭对称向外扩展。图 4.34(d)清楚地显示了中心化变换不仅有助于可视化,而且像前几章和本章所述的,也能简化滤波。

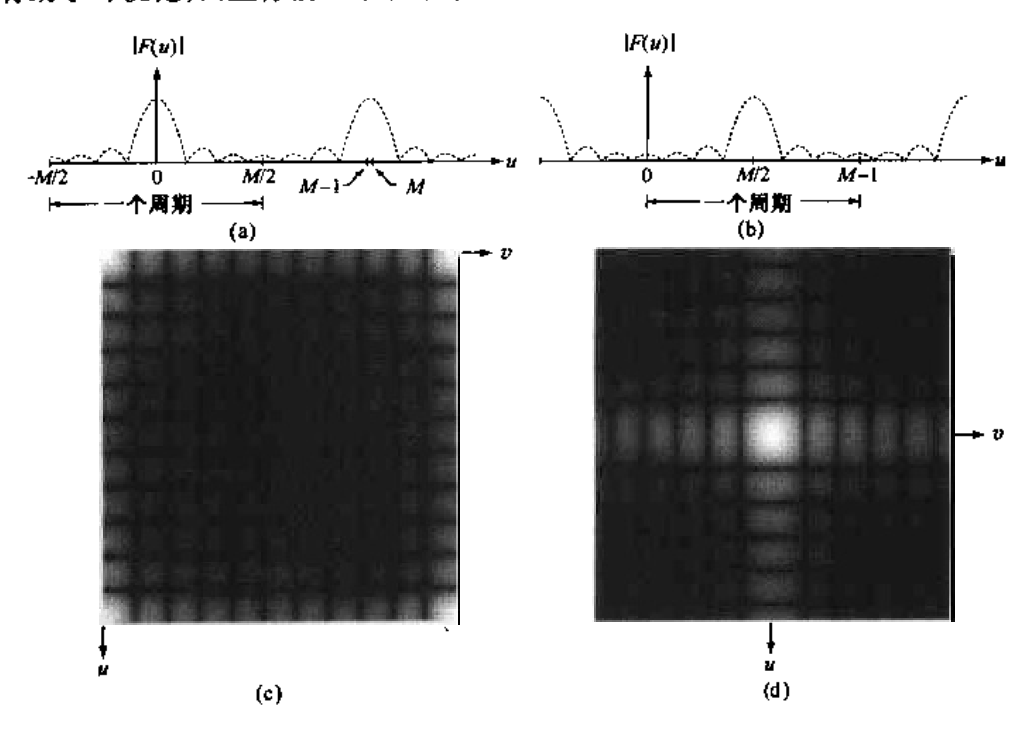


图 4.34 (a)在区间[0, M - 1]中显示半周期背靠背的傅里叶谱,(b)在相同区间内移动频谱以显示全周期,(c)一幅图像的傅里叶谱,显示了与(a)一样的背靠背特性,但是二维的,(d)中心化的傅里叶谱

可分性

式(4.2.16)的离散傅里叶变换可以用可分离的形式表示:

$$F(u,v)=\frac{1}{M}\sum_{x=0}^{M-1}e^{-j2\pi ux/M}\frac{1}{N}\sum_{y=0}^{N-1}f(x,y)e^{-j2\pi vy/N}$$

$$=\frac{1}{M}\sum_{x=0}^{M-1}F(x,v)e^{-j2\pi ux/M} \tag{4.6.14}$$

这里

$$F(x,v)=\frac{1}{N}\sum_{y=0}^{N-1}f(x,y)e^{-j2\pi vy/N} \tag{4.6.15}$$

对于每个 x 值,当 $v=0,1,2,\cdots,N-1$ 时,该等式是完整的一维傅里叶变换。换言之,$F(x,v)$ 是沿着 $f(x,y)$ 的一行所进行的傅里叶变换。当 x 由 0 变为 $M-1$ 时,沿着 $f(x,y)$ 的所有行计算傅里叶变换。然而频率变量 u 仍然保持不变。为完成二维变换,必须在式(4.6.14)中将 u 值从 0 变到 $M-1$。很明显,这涉及沿 $F(x,v)$ 的每一列计算一维变换。这是一个重要结论。它告诉我们可以通过先沿输入图像的每一行计算一维变换,然后沿中间结果的每一列再

计算一维变换的方法来求二维变换。颠倒次序后(先列后行)结论同样成立。该步骤概括于图 4.35 中。

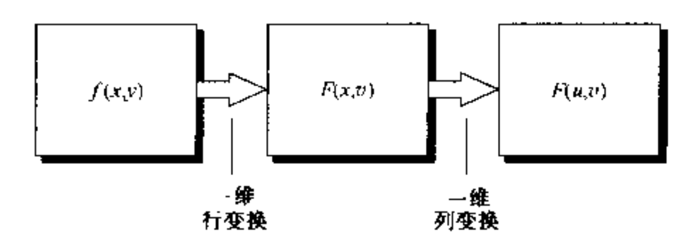


图 4.35　顺序进行一维变换计算二维傅里叶变换

一个相似的过程适用于计算二维傅里叶反变换。先沿 $F(u,v)$ 的每一行计算一维反变换,再沿中间结果的每一列计算一维反变换。正如下面章节所示,可用一维前向傅里叶变换算法实现反变换。

4.6.2　用前向变换算法计算傅里叶反变换

如前面一节所述,二维傅里叶变换可以通过应用一维变换实现。在 4.2.1 节定义了一维傅里叶变换对。为方便,重复这两个公式:

$$F(u) = \frac{1}{M}\sum_{x=0}^{M-1} f(x)e^{-j2\pi ux/M} \tag{4.6.16}$$

$u=0,1,2,\cdots,M-1$,且

$$f(x) = \sum_{u=0}^{M-1} F(u)e^{j2\pi ux/M} \tag{4.6.17}$$

$x=0,1,2,\cdots,M-1$。式(4.6.17)取复共轭,并用 M 同时除以两边,得到:

$$\frac{1}{M}f^*(x) = \frac{1}{M}\sum_{u=0}^{M-1} F^*(u)e^{-j2\pi ux/M} \tag{4.6.18}$$

与式(4.6.16)比较这一结果显示出式(4.6.18)的右边是前向傅里叶变换的形式。因此,将 $F^*(u)$ 输入到为计算前向变换而设计的算法中,得到值 $f^*(x)/M$。取复共轭,再乘以 M 便产生所要的反变换 $f(x)$。对双变量有类似的分析:

$$\frac{1}{MN}f^*(x,y) = \frac{1}{MN}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1} F^*(u,v)e^{-j2\pi(ux/M+vy/N)} \tag{4.6.19}$$

这就是二维前向傅里叶变换的形式。如果 $f(x)$ 或 $f(x,y)$ 是实函数(如一幅图像),式(4.6.19)左边取复共轭就没有必要了。只简单地取出结果的实部而忽略寄生的复数项,在大部分傅里叶变换的计算中是很典型的。

当开发出的技术用于得到变换时,通过相继的一维变换计算二维变换经常是混乱的根源。记住前一节列出的步骤,避免被式(4.6.18)误导。换句话说,当一维算法被用于计算二维反变换时,在处理每一行或一列后不必计算共轭。取而代之的,函数 $F^*(u,v)$ 被当做 $f(x,y)$,用在前向二维变换步骤中,二维变换步骤总结在图 4.35 中。结果中的复共轭(或可用的实部)乘以 $M*N$,得到相对应的正确 $f(x,y)$。这里强调,上述关于常数 M 和 N 的讨论是基于离散傅里叶变换的定义,即所有的常量和前向变换有关。如 4.2.1 节和 4.2.2 节指出的那样,通常会遇到常量以不同方式分布在前向变换和反变换之间的情况。因此,为避免减去一个缩放因子,当计算反变换时,如果以与本书不同的方式分布这些常量,要特别注意常量的放置。

4.6.3 更多的关于周期性的讨论:必要的铺垫

4.2.4 节解释了,基于卷积理论,频率域的乘法相当于空间域的卷积,反之亦然。当处理离散变量和傅里叶变换时,要记住不同函数所包含的周期性(4.6.1 节)。虽然可能不太直观,但周期性是定义离散傅里叶变换对时产生的数学副产品。周期性是处理操作的一部分,不应忽视。

图 4.36 列举了周期性的重要性。图的左边一列是用式(4.2.30)的一维形式计算的卷积:

$$f(x) * h(x) = \frac{1}{M}\sum_{m=0}^{M-1} f(m)h(x-m) \tag{4.6.20}$$

在此,还要利用这一机会更详细地解释卷积运算。为简化表示,简单的数字将代替那些表示函数长度和高度的通用符号。图 4.36(a)和(b)是两个要进行卷积的函数。每个函数包含 400 个点。卷积的第一步是将一个函数关于原点进行镜像映射(倒转),在本例情况下,对第二个函数进行,在图 4.36(c)中以 $h(-m)$示出。下一步是将 $h(-m)$滑过 $f(m)$。这要增加一个常数 x 到 $h(-m)$,即变成 $h(x-m)$,如图 4.36(d)所示。注意只有一个置换值。在第一次遇到时,这个简单步骤通常是引起混乱的根源。而这恰好是卷积计算的全部关键。换言之,为了执行卷积,倒转了一个函数,并将它滑过另一个函数。在每一个置换点(的每一个值)都要计算式(4.6.20)的全部总和。这个总和不比在给定位移处 f 和 h 乘积的和更大。位移 x 的范围为 h 完全滑过f需要的所有值。图 4.36(e)显示了 h 完全滑过f后的结果,并在 x 的每个点计算式(4.6.20)。在此例中,为使 $h(x-m)$完全滑过 f,x 值的范围是从 0 到 799。这幅图是两个函数的卷积。要清楚地记住卷积中的变量是 x。

从 4.2 节[见式(4.2.31)]介绍的卷积理论可知,利用式(4.6.20),由 $F(u)H(u)$的傅里叶反变换能得到同样的准确结果。但是,从本章前面对周期性的讨论又知离散傅里叶变换自动地将输入函数周期化。换言之,采用 DFT 允许在频率域进行卷积计算,但函数必须看做周期性的,且周期等于函数的长度。

可以通过图 4.36 右边一列考察这种隐含的周期性。图 4.36(f)同图 4.36(a)一样,但同样的函数在两个方向上周期性地无限扩展(扩展部分用虚线表示)。从图 4.36(g)到图 4.36(i)同样应用该扩展。现在,可以通过将 $h(x-m)$滑过 $f(m)$进行卷积。如前面一样,变化 x 完成滑动。然而,$h(x-m)$的周期性扩展产生了图4.36左边的计算中所没有的值。例如,在图 4.36(i)中,当 $x=0$ 时,看到 $h(x-m)$右侧第一个扩展周期的一部分进入图 4.36(f)中所示的 $f(m)$(从原点开始)的一部分。当 $h(x-m)$向右滑动时,在 $f(m)$中的那部分开始向右侧移出,但被 $h(x-m)$左侧相同部分所取代。这引起卷积产生一个常量值,如图 4.36(j)所示的[0,100]的一段。从 100 到 400 的一段是正确的,但周期性是周而复始的,这样就引起卷积函数尾部的一部分丢失,由图 4.36(j)和图 4.36(e)实线部分的比较可以看出这一点。

在频率域,该过程需要计算图 4.36(a)和(b)中函数的傅里叶变换。根据卷积理论,两个变换要相乘,再计算傅里叶反变换。结果包含 400 个点的卷积,如图 4.36(j)的实线部分所示。简单的解释表明当使用傅里叶变换得出卷积函数时,错误地处理周期性将得到错误的结论。结果,在开头有错误数据,结尾将丢失数据。

问题的解决办法很简单。假设 f 和 h 分别由 A 个和 B 个点组成。对两个函数同时添加

零，以使它们具有相同的周期，表示为 P。这个过程产生扩展的或延拓的函数，如下所示：

$$f_e(x) = \begin{cases} f(x) & 0 \leqslant x \leqslant A-1 \\ 0 & A \leqslant x \leqslant P \end{cases} \tag{4.6.21}$$

和

$$g_e(x) = \begin{cases} g(x) & 0 \leqslant x \leqslant B-1 \\ 0 & B \leqslant x \leqslant P \end{cases} \tag{4.6.22}$$

可以看出(Brigham[1988])，除非选择 $P \geqslant A+B-1$，否则卷积的独立周期将会混叠。已经在图 4.36中看到了这种现象的结果，这通常归于缠绕误差。若 $P = A+B-1$，周期便会邻接起来。若 $P > A+B-1$，周期将会是分隔开的，分隔的程度等于 P 与 $A+B-1$ 的差。

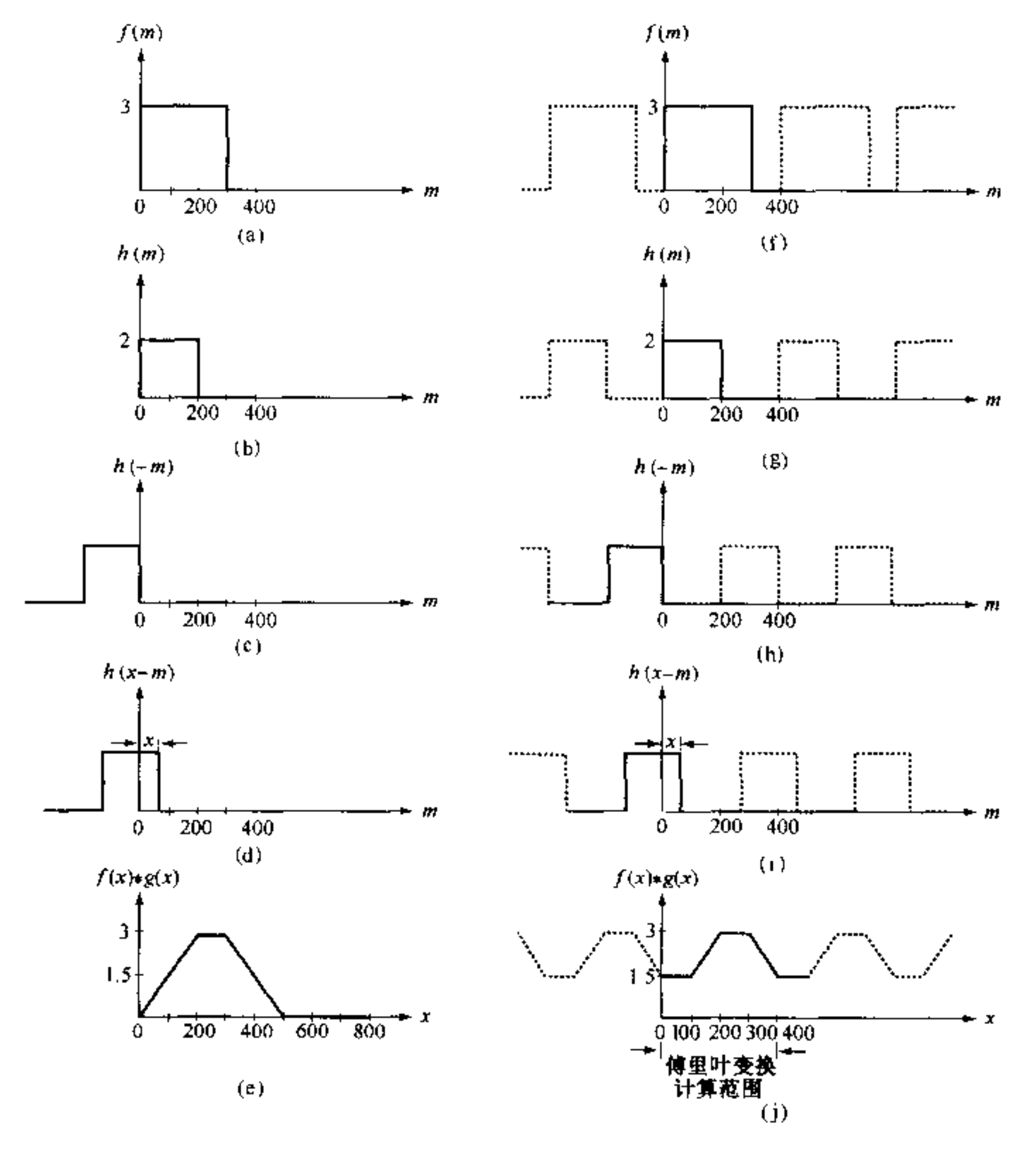


图 4.36　左边：两个离散函数的卷积。右边：相同函数的卷积，考虑 DFT 周期性的应用。注意在(j)中邻近周期的数据如何混淆卷积结果

扩展图 4.36(a)和(b)的函数后得到的结果示于图 4.37(a)和(b)。此例中，选择 $P = A + B - 1(799)$，即可知卷积周期是相邻的。遵循与前面的解释相同的过程，得到如图4.37(e)所示

的卷积函数。该结果的一个周期与图 4.36(e)相同,是正确的。这样,如果要在频率域计算卷积,应该:(1)得到两个扩展序列的傅里叶变换(每个序列有 800 个点);(2)将两个变换相乘;(3)计算傅里叶反变换。结果便得到正确的 800 个点的卷积函数,见图 4.37(e)中周期加重的部分。

这些概念扩展到二维函数时遵循了相同的前提。假设有 $f(x,y)$ 和 $h(x,y)$ 两幅图像,大小分别为 $A \times B$ 和 $C \times D$。如同一维情况,这些行列必须假定在 x 方向上有相同的周期 P,在 y 方向上有相同的周期 Q。二维卷积的混叠可由选择如下周期避免:

$$P \geqslant A + C - 1 \tag{4.6.23}$$

$$Q \geqslant B + D - 1 \tag{4.6.24}$$

扩展 $f(x,y)$ 和 $h(x,y)$ 形成如下周期性序列:

$$f_e(x,y) = \begin{cases} f(x,y) & 0 \leqslant x \leqslant A-1 \text{ 且 } 0 \leqslant y \leqslant B-1 \\ 0 & A \leqslant x \leqslant P \text{ 或 } B \leqslant y \leqslant Q \end{cases} \tag{4.6.25}$$

$$h_e(x,y) = \begin{cases} h(x,y) & 0 \leqslant x \leqslant C-1 \text{ 且 } 0 \leqslant y \leqslant D-1 \\ 0 & C \leqslant x \leqslant P \text{ 或 } D \leqslant y \leqslant Q \end{cases} \tag{4.6.26}$$

填充的问题是对滤波中心化。当执行本章讨论过的任何频率域滤波时,用要处理的图像变换乘以滤波器传递函数。由卷积定理可知,这与空间域滤波器表达式和图像相卷积的效果是相同的。这样,如果正确的填充不能实现,结果便会是错误的,如图 4.38 所示。为了简化图例,假设 f 和 h 是方形的,且大小相同,这里 h 是滤波器 $H(u,v)$ 的傅里叶反变换。图 4.38(a)显示了图像没有延拓时得到的滤波结果。这通常是由于没有对一幅输入图像进行延拓就进行傅里叶变换,然后又乘上同样大小的滤波函数(也没有延拓),计算傅里叶反变换。结果就是与输入图像相同的大小为 $A \times B$ 的图像,如图 4.38(a)左上象限所示。如同一维情况,图像前面边沿(阴影部分)由于周期性而引入了错误数据,而在尾部边沿将丢失数据。如图 4.38(b)所示,通过对输入图像和滤波函数进行合适的延拓,将得到正确的、大小为 $P \times Q$ 的过滤图像,如图 4.38(c)所示。这幅图像在两个坐标方向上是原始图像的两倍大小,有原始图像 4 倍数量的像素点。但是,不久就要看到,感兴趣的区域通常出现在大图像上。

重要的是注意到,刚才描述的处理方法需要频率域滤波函数的反变换,用 0 延拓,然后再进行正变换。所有过滤的其他方面已在 4.2.3 节描述过了。同样要注意滤波器的傅里叶反变换有实部和虚部。虽然滤波器的虚部在本书讨论中通常比实部小许多个数量级,但实际一般不能在傅里叶计算的中间过程中忽略虚部。这样,在通过前向变换的计算产生频率域滤波器延拓前,实部和虚部都将延拓。

图 4.39 显示了用来生成图 4.12(c)的理想低通滤波器延拓的空间域表示(只显示实部)。黑色显示延拓"0"的区域。选择理想低通滤波器的例子是因为它在空间域中有着很大部分的可见"结构"。所用的延拓是所要求的最小尺寸,当图像和滤波器是大小相同的方形时,在两个维度上都简单地倍乘原始图像大小。

图 4.40 所示的是用刚刚讨论的方法延拓函数滤波的结果。很容易看出,用图 4.12(a)的延拓怎样卷绕图 4.39 所示的滤波器生成图 4.40。在这种情况下,很明显,结果图像的 3 个区域包含了无用的信息,因此裁剪回原始图像的大小可以得到想要的过滤结果。可以确定切去由于延拓产生的图像没有混叠错误。

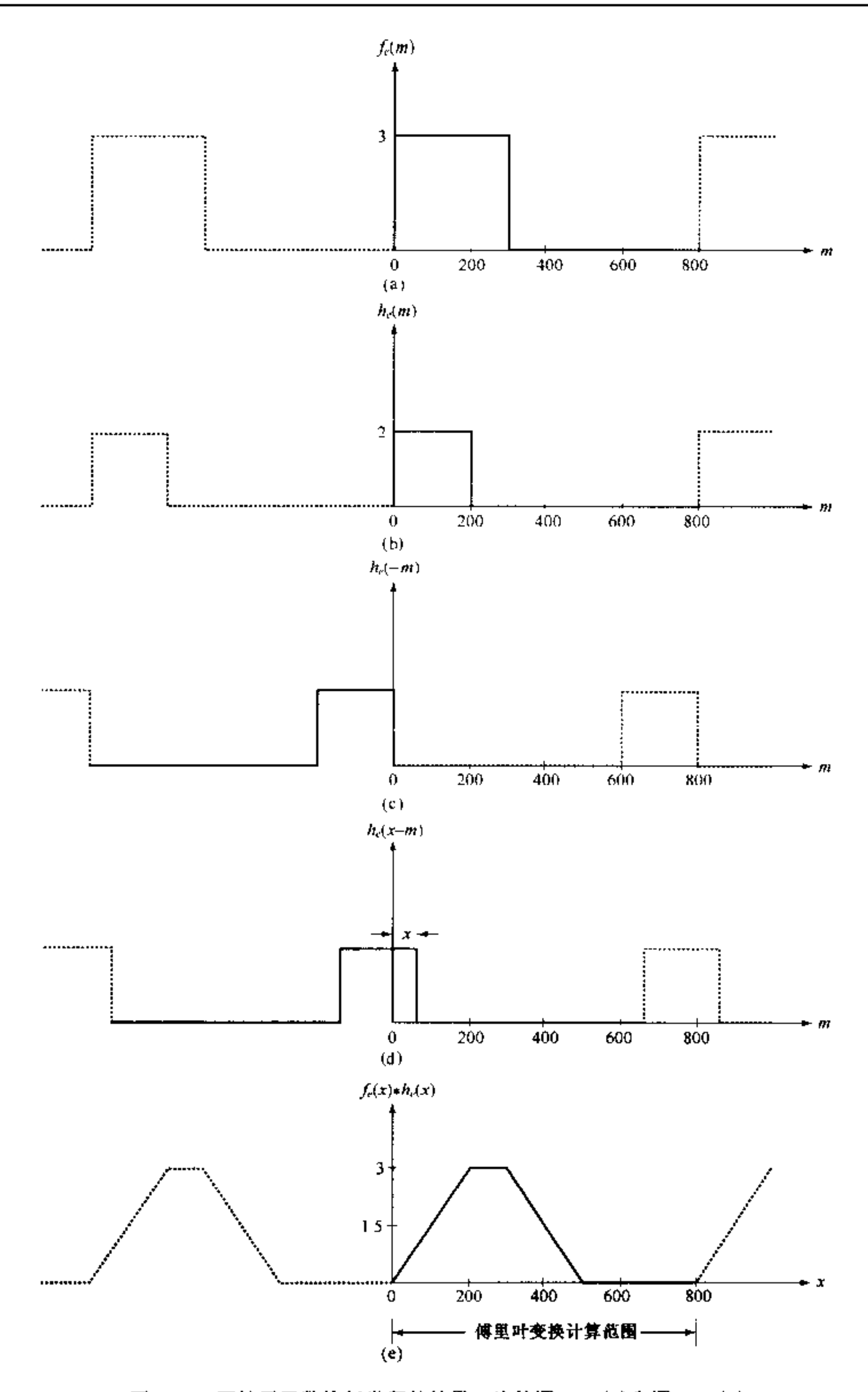


图 4.37　用扩展函数执行卷积的结果。比较图 4.37(e)和图 4.36(e)

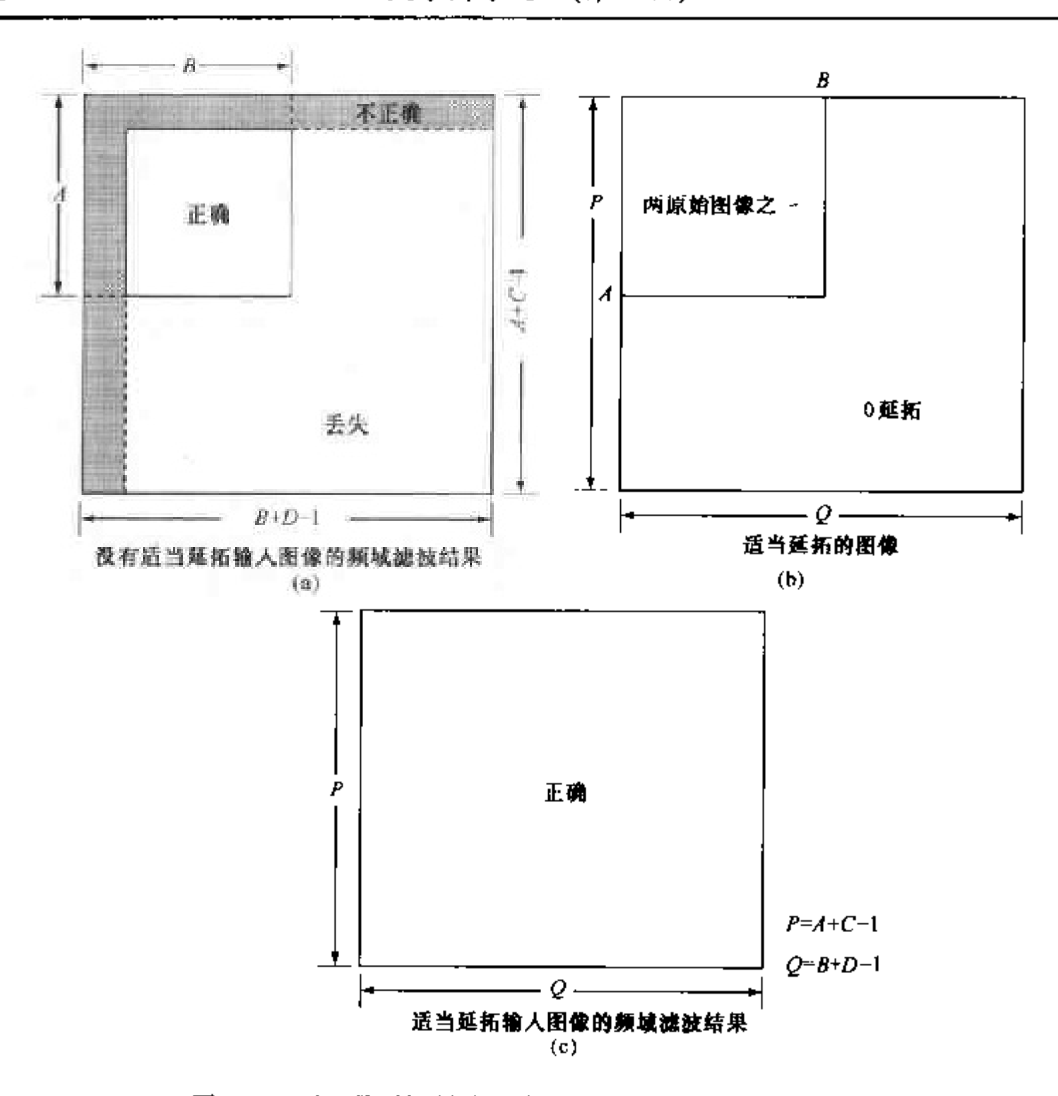


图 4.38 对函数延拓所需的说明,(a) 没有延拓,执行二维卷积的结果,(b)合适的函数延拓,(c)正确的卷积结果

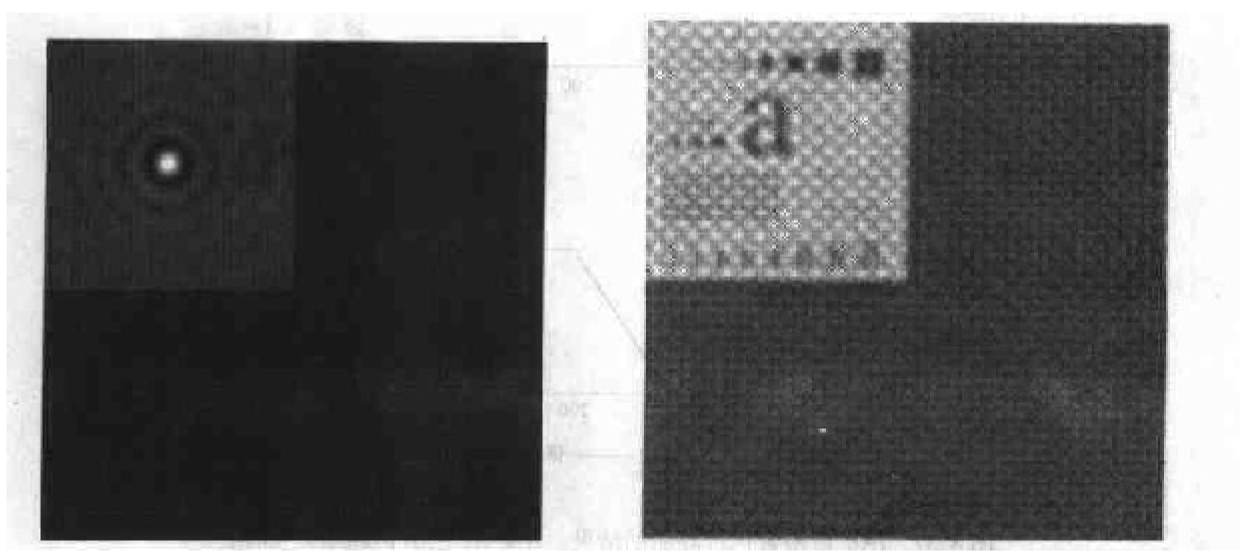


图 4.39 在空间域延拓的低通滤波器(仅显示实部)

图 4.40 用延拓滤波的结果。图像通常被裁剪到原来的尺寸,因为超过图像边界没有可见信息

4.6.4 卷积和相关性理论

在 4.2.4 节介绍了卷积,在 4.6.3 节讨论了实现卷积的其他细节。在这里简要地重复它是为了易于与相关这一类似过程进行比较。大小为 $M\times N$ 的两个函数$f(x,y)$和 $h(x,y)$的离散卷积表示为 $f(x,y)*h(x,y)$,由下面的表达式所定义:

$$f(x,y)*h(x,y)=\frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}f(m,n)h(x-m,y-n) \tag{4.6.27}$$

从 4.2.4 节的讨论可知,卷积理论由两个函数和它们的傅里叶变换间的下述关系组成:

$$f(x,y)*h(x,y)\Leftrightarrow F(u,v)H(u,v) \tag{4.6.28}$$

$$f(x,y)h(x,y)\Leftrightarrow F(u,v)*H(u,v) \tag{4.6.29}$$

两个函数 $f(x,y)$和 $h(x,y)$的相关性定义如下:

$$f(x,y)\circ h(x,y)=\frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}f^*(m,n)h(x+m,y+n) \tag{4.6.30}$$

这里,f^* 表示 f 的复共轭。一般处理的是实函数(图像),$f^*=f$。除了复共轭及和式第二项中的负号换为正号外,相关函数与式(4.6.27)给出的卷积函数有着相同的形式。这意味着 h 并不是关于原点的镜像。实现卷积的每一步都与相关相同,包括需要延拓。

给出卷积与相关的相似性后,相关理论类似于卷积理论就不奇怪了。让 $F(u,v)$和 $H(u,v)$分别表示 $f(x,y)$和 $h(x,y)$的傅里叶变换。卷积理论有一半讲的是空间域的相关[$f(x,y)\circ h(x,y)$]和频率域的乘积[$F^*(u,v)H(u,v)$],它们组成了一个傅里叶变换对。这一结论形式上如下所示:

$$f(x,y)\circ h(x,y)\Leftrightarrow F^*(u,v)H(v,v) \tag{4.6.31}$$

它指出空间域的相关性可由 $F^*(u,v)H(u,v)$的傅里叶反变换得到,F^* 是 F 的复共轭。一个类似的结果是:频率域的相关简化为空间域的相乘。即:

$$f^*(x,y)h(x,y)\Leftrightarrow F(u,v)\circ H(v,v) \tag{4.6.32}$$

这两个结论构成了相关理论。假定所有的函数被正确地延拓扩展了。

现在,我们知道了卷积是空间域过滤和频率域过滤之间的纽带。相关的重要用途在于匹配。在匹配中,$f(x,y)$是一幅包含物体或区域的图像。如果想要确定 f 是否包含有感兴趣的物体或区域,让$h(x,y)$作为那个物体或区域(通常称该图像为模板)。那么,如果匹配,两个函数的相关值会在 h 找到f 中相应点的位置上达到最大。前期的处理,如缩放和对齐,在多数的实际应用中都是必需的,但大批量的处理工作是进行相关运算。

最后,指出术语"互相关"常常用来代替术语"相关",为的是指明进行相关运算的图像是不同的。这是相对自相关而言的,在自相关中,两幅图像是相同的。在后面的例子中,我们将讨论自相关理论,它由式(4.6.31)直接得到:

$$f(x,y)\circ f(x,y)\Leftrightarrow |F(u,v)|^2 \tag{4.6.33}$$

在右边,利用了"复数和它的复共轭的乘积是复数模的平方"这一结论。也就是说,这一结果表明空间域自相关的傅里叶变换是式(4.2.20)定义的功率谱。类似地,

$$|f(x,y)|^2\Leftrightarrow F(u,v)\circ F(u,v) \tag{4.6.34}$$

例 4.11 图像相关

图 4.41 显示了一个图像延拓和相关的简单例子。图 4.41(a)是该图像,图 4.41(b)是模板。图像和模板的大小分别是 256×256 和 38×42 像素。此例中,$A=B=256$,$C=38$ 和 $D=42$。这就给出了扩展函数的最小值:$P=A+C-1=293$ 和 $Q=B+D-1=297$。

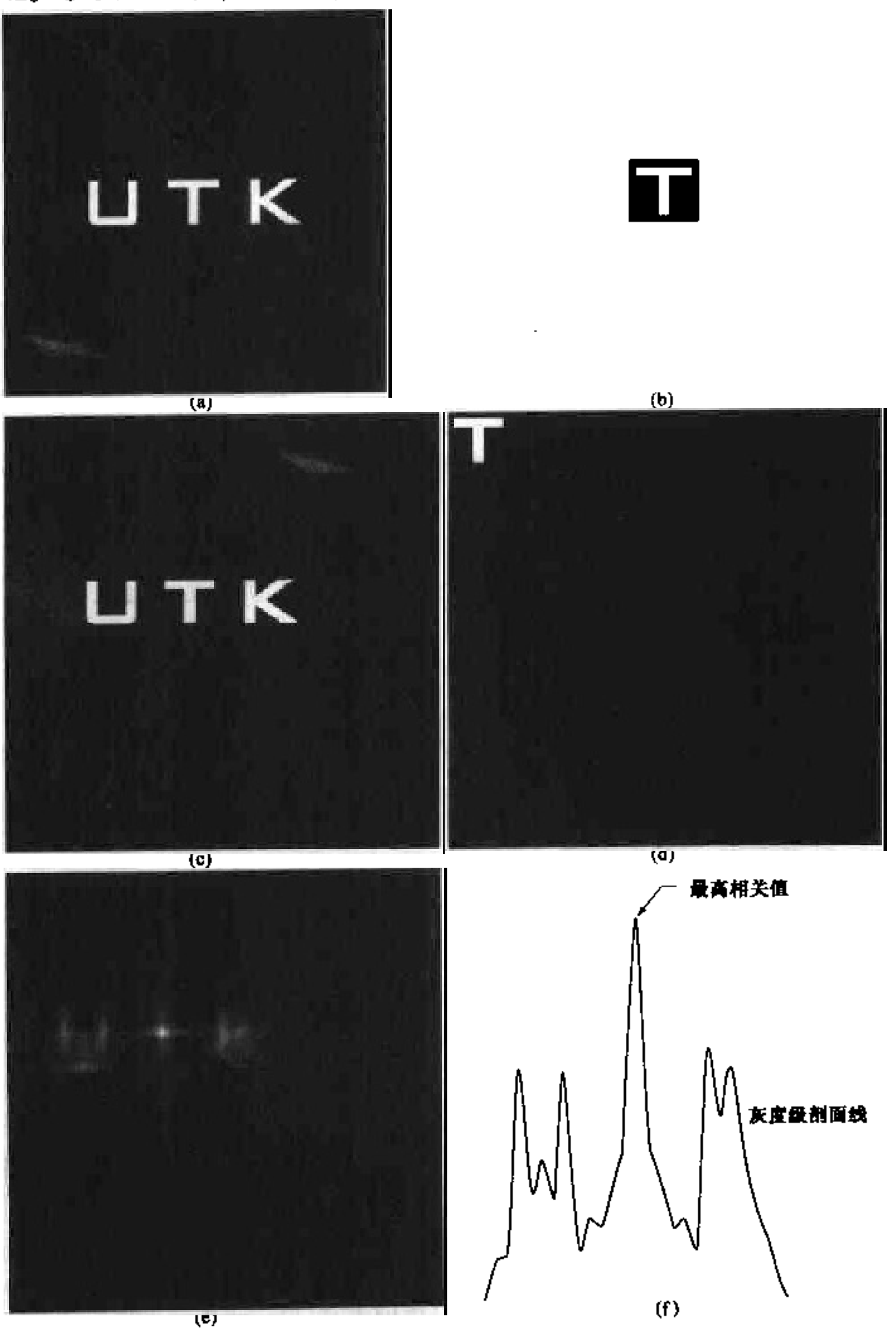


图 4.41 (a)图像,(b)模板,(c)和(d)延拓图像,(e)以图像显示的相关函数,(f)通过(e)中最高值的水平剖面线。所示的点是最佳匹配点

这里选择 298×298 的相等延拓尺度。延拓后的图像如图 4.41(c)和(d)所示。两延拓图像的空间域相关以图像形式显示于图 4.41(e)。正如式(4.6.31)指出的那样，产生相关函数需要计算延拓图像的变换，从它们中选出一个来取复共轭(我们选择模板)，将两个变换相乘再求傅里叶反变换。留下一个练习给读者(习题 4.23)，讨论如果取的是另外一个变换的共轭，图 4.41(e)将会是什么样子。

如预料的那样，从图 4.41(e)看出相关函数最大值正好出现在图像中“T”的顶部——模板所在的点上。如同卷积，重要的是记住相关函数中的变量在空间域中是可移动的。例如，图 4.41(e)的左上角相当于一个函数关于另一个函数的零位移。图 4.41(e)中每个像素点的值都是相关函数在一个位移位置的值，即是式(4.6.30)中 (x,y) 数对的一个特殊值。还注意到，相关函数与延拓图像有着相同的尺寸。最后，图 4.41(f)显示了通过图4.41(e)最大值的水平灰度剖面图。这张图简单地确认了相关函数的最高峰位于模板和图像发生最佳匹配的点上。

4.6.5　二维傅里叶变换性质总结

表 4.1 中概括了本章讨论的所有傅里叶变换的性质。脚注标识了需要延拓函数的项，以避免不正确的结果。如前面所述，双箭头用于标明傅里叶变换对的表达形式。即双箭头右边的表达式由对左边表达式进行前向傅里叶变换得到，左侧表达式由右侧表达式进行傅里叶反变换得到。

表 4.1　二维傅里叶变换重要性质总结

性　质	表 达 式
傅里叶变换	$F(u,v)=\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)e^{-j2\pi(ux/M+vy/N)}$
傅里叶反变换	$f(x,y)=\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}F(u,v)e^{j2\pi(ux/M+vy/N)}$
极坐标表示	$F(u,v)=\lvert F(u,v)\rvert e^{-j\phi(u,v)}$
谱	$\lvert F(u,v)\rvert=[R^2(u,v)+I^2(u,v)]^{1/2}$，$R=\mathrm{Real}(F)$ $I=\mathrm{Imag}(F)$
相位角	$\phi(u,v)=\arctan\left[\frac{I(u,v)}{R(u,v)}\right]$
功率谱	$P(u,v)=\lvert F(u,v)\rvert^2$
均值	$\bar{f}(x,y)=F(0,0)=\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)$
平移	$f(x,y)e^{j2\pi(u_0x/M+v_0y/N)}\Leftrightarrow F(u-u_0,v-v_0)$ $f(x-x_0,y-y_0)\Leftrightarrow F(u,v)e^{-j2\pi(ux_0/M+vy_0/N)}$ 当 $x_0=u_0=M/2$ 和 $y_0=v_0=N/2$ 时， $f(x,y)(-1)^{x+y}\Leftrightarrow F(u-M/2,v-N/2)$ $f(x-M/2,y-N/2)\Leftrightarrow F(u,v)(-1)^{u+v}$
共轭对称	$F(u,v)=F^*(-u,-v)$ $\lvert F(u,v)\rvert=\lvert F(-u,-v)\rvert$

(续表)

性　质	表 达 式
微分	$\frac{\partial^n f(x,y)}{\partial x^n} \Leftrightarrow (ju)^n F(u,v)$ $(-jx)^n f(x,y) \Leftrightarrow \frac{\partial^n F(u,v)}{\partial u^n}$
拉普拉斯	$\nabla^2 f(x,y) \Leftrightarrow -(u^2+v^2)F(u,v)$
分配性	$\Im[f_1(x,y)+f_2(x,y)] = \Im[f_1(x,y)] + \Im[f_2(x,y)]$ $\Im[f_1(x,y)\cdot f_2(x,y)] \neq \Im[f_1(x,y)]\cdot\Im[f_2(x,y)]$
比例性	$af(x,y) \Leftrightarrow aF(u,v), f(ax,by) \Leftrightarrow \frac{1}{\lvert ab \rvert}F(u/a,v/b)$
旋转	$x=r\cos\theta \quad y=r\sin\theta \quad u=\omega\cos\varphi \quad v=\omega\sin\varphi$ $f(r,\theta+\theta_0) \Leftrightarrow F(\omega,\varphi+\theta_0)$
周期性	$F(u,v)=F(u+M,v)=F(u,v+N)=F(u+M,v+N)$ $f(x,y)=f(x+M,y)=f(x,y+N)=f(x+M,y+N)$
可分性	见式(4.6.14)和式(4.6.15)。可分性意指可以首先沿着图像的每一行计算一维变换,然后沿着这一中间结果的每一列计算一维变换,以此计算二维变换,得到相同的结果。
使用前向变换算法的傅里叶反变换的计算	$\frac{1}{MN}f^*(x,y) = \frac{1}{MN}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}F^*(u,v)e^{-j2\pi(ux/M+vy/N)}$ 这一等式指出输入一个函数 $F^*(u,v)$ 到用于计算前向变换的算法(前面等式的右边)得到 $f^*(x,y)/MN$。取复共轭并用 MN 乘以这一结果得到希望的反变换。
卷积*	$f(x,y)*h(x,y) = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}f(m,n)h(x-m,y-n)$
相关*	$f(x,y)\circ h(x,y) = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}f^*(m,n)h(x+m,y+n)$
卷积定理*	$f(x,y)*h(x,y) \Leftrightarrow F(u,v)H(u,v)$; $f(x,y)h(x,y) \Leftrightarrow F(u,v)*H(u,v)$
相关定理*	$f(x,y)\circ h(x,y) \Leftrightarrow F^*(u,v)H(u,v)$; $f^*(x,y)h(x,y) \Leftrightarrow F(u,v)\circ H(u,v)$
某些有用的 FT 对:	
脉冲	$\delta(x,y) \Leftrightarrow 1$
高斯	$A\sqrt{2\pi}\sigma e^{-2\pi^2\sigma^2(x^2+y^2)} \Leftrightarrow Ae^{-(u^2+v^2)/2\sigma^2}$
矩形	$\mathrm{rect}[a,b] \Leftrightarrow ab\frac{\sin(\pi ua)}{(\pi ua)}\frac{\sin(\pi vb)}{(\pi ua)}e^{-j\pi(ua+vb)}$
余弦	$\cos(2\pi u_0x+2\pi v_0y) \Leftrightarrow \frac{1}{2}[\delta(u+u_0,v+v_0)+\delta(u-u_0,v-v_0)]$
正弦	$\sin(2\pi u_0x+2\pi v_0y) \Leftrightarrow j\frac{1}{2}[\delta(u+u_0,v+v_0)-\delta(u-u_0,v-v_0)]$

*假定函数已经用 0 延拓扩展了。

4.6.6 快速傅里叶变换

如 4.1 节指出的那样,离散傅里叶变换成为信号处理的一种基础工具的一个主要原因是快速傅里叶变换(FFT)的发展。用式(4.2.5)计算 M 点一维傅里叶变换直接要求 M^2 次的乘法/加法运算。快速傅里叶变换完成同样任务则只需进行 $M\log_2 M$ 次的运算。例如当 $M=1\,024$ 时,用笨拙的方法大约需要 10^6 次操作,而快速傅里叶变换只需要约 10^4 次操作即可。这是

100:1的运算优势。如果还不能看到这种优势的重要性所在,试想一下能够在一年内完成的规定工作反之却用 100 年完成!这就是理论的可能性与现实的不可行性。所要解决的问题越大,用快速傅里叶变换计算的优势也就越大。例如,若 $M = 8\ 192(2^{13})$,计算的优势就增加到 600:1。这样诱人的数字成为想要更多地了解快速傅里叶变换算法的巨大动力。这一节先了解导出快速傅里叶变换的离散傅里叶变换的基本分解如何推导。重点在单变量的快速傅里叶变换。如 4.6.1 节所指出的那样,二维傅里叶变换能通过相继的一维变换算法得到。

本节中推导的快速傅里叶变换算法基于"逐次倍乘法"。为表达方便,以下列形式表示式(4.2.5):

$$F(u) = \frac{1}{M}\sum_{x=0}^{M-1} f(x) W_M^{ux} \tag{4.6.35}$$

这里

$$W_M = e^{-j2\pi/M} \tag{4.6.36}$$

假设 M 的形式是:

$$M = 2^n \tag{4.6.37}$$

n 为正整数。因此,M 可以表示为:

$$M = 2K \tag{4.6.38}$$

K 也是一个正整数。将式(4.6.38)代入式(4.6.35)得到:

$$\begin{aligned} F(u) &= \frac{1}{2K}\sum_{x=0}^{2K-1} f(x) W_{2K}^{ux} \\ &= \frac{1}{2}\left[\frac{1}{K}\sum_{x=0}^{K-1} f(2x) W_{2K}^{u(2x)} + \frac{1}{K}\sum_{x=0}^{K-1} f(2x+1) W_{2K}^{u(2x+1)}\right] \end{aligned} \tag{4.6.39}$$

然而,用式(4.6.36)可得 $W_{2K}^{2ux} = W_K^{ux}$。因此,式(4.6.39)可表示成:

$$F(u) = \frac{1}{2}\left[\frac{1}{K}\sum_{x=0}^{K-1} f(2x) W_{2K}^{ux} + \frac{1}{K}\sum_{x=0}^{K-1} f(2x+1) W_K^{ux} W_{2K}^{u}\right] \tag{4.6.40}$$

定义:

$$F_{\text{even}}(u) = \frac{1}{K}\sum_{x=0}^{K-1} f(2x) W_K^{ux} \tag{4.6.41}$$

$u = 0, 1, 2, \cdots, K-1$,且:

$$F_{\text{odd}}(u) = \frac{1}{K}\sum_{x=0}^{K-1} f(2x+1) W_K^{ux} \tag{4.6.42}$$

$u = 0, 1, 2, \cdots, K-1$,式(4.6.40)就变为:

$$F(u) = \frac{1}{2}\left[F_{\text{even}}(u) + F_{\text{odd}}(u) W_{2K}^{u}\right] \tag{4.6.43}$$

同样,因为 $W_M^{u+M} = W_M^u$ 和 $W_{2M}^{u+M} = -W_{2M}^u$,式(4.6.41)通过式(4.6.43)得到:

$$F(u+K) = \frac{1}{2}\left[F_{\text{even}}(u) - F_{\text{odd}}(u) W_{2K}^{u}\right] \tag{4.6.44}$$

仔细分析式(4.6.41)至式(4.6.44)可得出这些表达式的一些有趣性质。一个 M 点变换可以通过把原始表达式分成两部分来计算,如式(4.6.43)和式(4.6.44)所示。计算 $F(u)$ 的前半部分要对式(4.6.41)和式(4.6.42)给出的两个 $M/2$ 点变换进行计算。$F_{\text{even}}(u)$ 和 $F_{\text{odd}}(u)$ 的计算结果被代入式(4.6.43)中得到 $F(u)$,$u = 0, 1, 2, \cdots, (M/2-1)$。另外一半可直接从

式(4.6.44)得到,而无需另外的变换计算。

为了检验这个计算步骤可否实现,令 $m(n)$ 和 $a(n)$ 分别代表实现算法所要求的复数乘法和加法的次数。如前所述,样点的数量为 2^n, n 为正整数,先假设 $n=1$。两点变换要计算 $F(0)$;然后,$F(1)$ 由式(4.6.44)得到。为得到 $F(0)$,首先要计算 F_{even} 和 F_{odd}。此例中 $K=1$ 且式(4.6.41)和式(4.6.42)是一点变换。因为单点傅里叶变换即为取样点自身,因而,得到 $F_{even}(0)$ 和 $F_{odd}(0)$ 不需要任何乘法和加法运算。$F_{odd}(0)$ 乘以 W_2^0,再使用一次加法,由式(4.6.43)得到 $F(0)$。然后,从式(4.6.44)得到 $F(1)$ 又用了一次加法(减法可以认为与加法相同)。当计算完 $F_{odd}(0)W_2^0$ 后,一次两点变换所需要的总运算次数由 $m(1)=1$ 次乘法和 $a(1)=2$ 次加法构成。

n 的下一个值为 2。根据前面的过程,一个 4 点变换可分成两部分。$F(u)$ 的前半部分要求对两个两点变换求值,如 $K=2$ 是式(4.6.41)和式(4.6.42)给出的。一个两点变换要求 $2m(1)$ 次乘法和 $2a(1)$ 次加法。因此,这两个等式的计算总共要求 $2m(1)$ 次乘法和 $2a(1)$ 次加法。为从式(4.6.43)得到 $F(0)$ 和 $F(1)$,进一步需要两次乘法和加法。因为 $F_{odd}(u)W_{2k}^u$ 在 $u=\{0,1\}$ 时已被计算,所以要得到 $F(2)$ 和 $F(3)$ 又要进行两次加法计算。这时总数是 $m(2)=2m(1)+2$ 和 $a(2)=2a(1)+4$。

当 $n=3$ 时,两个 4 点变换用于计算 $F_{even}(u)$ 和 $F_{odd}(u)$。它们需要 $2m(2)$ 次乘法和 $2a(2)$ 次加法。完成全部变换还要 4 次乘法和 8 次加法。总数为 $m(3)=2m(2)+4$ 和 $a(3)=2a(2)+8$。

对任意正整数 n,可以推导出完成快速傅里叶变换所需的乘法和加法次数的递归表达式:

$$m(n) = 2m(n-1) + 2^{n-1} \qquad n \geqslant 1 \tag{4.6.45}$$

和

$$a(n) = 2a(n-1) + 2^n \qquad n \geqslant 1 \tag{4.6.46}$$

这里,$m(0)=0$ 且 $a(0)=0$,因为单点变换无需任何加法和乘法运算。

式(4.6.41)至式(4.6.44)的实现构成了逐次倍乘的快速傅里叶变换算法。名字是由计算方法(即对于任意的 2 的整数次幂 M,一个两点变换的计算来自两个单点变换,一个 4 点变换来自两个两点变换,等等)得来的。作为练习留给读者(习题 4.25),说明:

$$m(n) = \frac{1}{2}M\log_2 M \tag{4.6.47}$$

和

$$a(n) = M\log_2 M \tag{4.6.48}$$

快速傅里叶变换优于直接用一维离散傅里叶变换的计算优势定义为:

$$C(M) = \frac{M^2}{M\log_2 M} = \frac{M}{\log_2 M} \tag{4.6.49}$$

因为假设 $M=2^n$,所以可以用 n 的形式来表示式(4.6.49):

$$C(n) = \frac{2^n}{n} \tag{4.6.50}$$

该函数的曲线参见图 4.42。如前所述,迅速增加的优势是 n 的函数。例如,当 $n=15$(32,768 点)时,FFT 对于 DFT 已有接近 2 200:1 的优势。这样,可以预料在同一台机器上,快速傅里叶

变换以比离散傅里叶变换快 2 200 倍的速度进行计算。

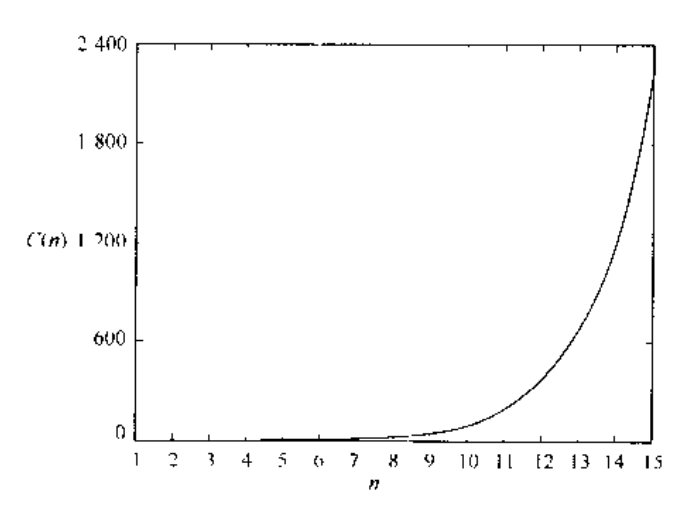


图 4.42　FFT 相对于直接实现一维 DFT 的计算优势。注意，该优势以 n 的函数快速增长

快速傅里叶变换优异的性质非常多，这里不再深入地讨论这个课题了（可参见 Brigham[1998]）。实际上，所有综合性的信号和图像处理软件包都已包含 FFT 的实现，也能够处理点数不为 2 的整数次幂的情况（以牺牲计算效率为代价）。免费的 FFT 程序已经可以从 Internet 得到。

4.6.7　关于滤波器设计的一些评论

本章讨论的所有滤波器都以公式形式来描述。为了使用这些滤波器，简单地对公式采样得到所希望的(u,v)值。这样就导出了滤波函数 $H(u,v)$。在所有的例子中，该函数先乘上输入图像的 DFT（中心化后），然后计算傅里叶反变换。本章中所有正向和反向的傅里叶变换都由 FFT 算法计算，计算采用图 4.35 和 4.2.6 节概括的步骤。

本章讨论的滤波方法严格围绕基本原理，重点在于尽可能清晰地解释频率域滤波的结果。可以将这里所用的滤波处理方法看做一个滤波器原型的基础。换言之，给出一个问题，为解决该问题，寻找某个滤波器，频率域处理方法对于实验、迅速而全面地控制滤波器参数是一个理想的工具。

一旦用于一个特殊应用的滤波器找到了，在空间域采用固件和/或硬件实现它通常是十分有趣的。Petrou 和 Bosdogianni[1999]提供了在二维频率域滤波器和相应的数字滤波器之间很好的纽带。关于二维数字滤波器的设计见 Lu 和 Antonion[1992]的论著。

小结

本章提供的资料与第 3 章的阐述结合起来是图像增强完整的基础内容。毫无疑问，读者现在已清楚地了解到，图像增强技术在实际应用中确实能创造出令人满意的结果。大多数工具都基于数学和统计学概念，但是，它们的用途是严格面向问题的。换言之，图像增强与其说是一门科学，其实更像一门艺术，一个正确增强图像的定义是高度主观化的。在第 5 章，将扩

展本章开发的一些数学概念,最后一章则转向图像复原领域。与增强不同,复原技术为的是改善图像,所以往往基于客观基础,而不是主观性准则。这样,图像复原技术比我们迄今为止所用的方法更结构化。

本章另一个主要目标是开发傅里叶变换。虽然这在图像增强一章中有所体现,但这种方法是非常通用的。在后续章节中,将看到离散傅里叶变换的不同应用。

参考资料

4.1 节的其他读物见 Hubbard[1998]的著述。由 Bracewell[2000],Bracewell[1995]所著的书籍是解释连续傅里叶变换和如何延伸到二维图像处理的入门书籍。这两本书以及 Lim[1990],Castleman[1996],Petrou 和 Bosdogianni[1999]及 Brigham[1988]的论著对 4.2 节大部分的内容提供了背景知识。

4.3 节和 4.4 节的附加读物可见 Castleman[1996],Pratt[1991]和 Hall[1997]的论著。滤波器实现的有效处理问题仍然是一个令人感兴趣的题目,如 Bakir 和 Reeves[2000]所述的例证。对于钝化模板和提升滤波可见 Schowengerdt[1983]的论著。同态滤波器(4.5 节)的资料基于 Stockham[1972]的论文;也可以参见 Oppenheim 和 Schafer[1975],Pitas 与 Venetsanopoulos[1990]的论著。Brinkman 等[1998]将钝化模板和同态滤波相结合用于增强磁共振图像。数字滤波器的生成(4.6.7 节)是基于本章所讨论的频率域公式,见 Lu 和 Antoniou[1992]以及 Petrou 与 Bosdogianni[1999]的著述。

如 4.1 节所述,快速傅里叶变换(FFT)的发现(4.6.6 小节)是离散傅里叶变换作为数字信号处理的基本工具而普及的重要里程碑。在 4.6.6 节对 FFT 的表述是基于 Cooley 和 Tuckey[1965]的论文,以及 Brigham[1988]的著作(他也讨论了几个 FFT 的实现,包括基数不为 2 的快速傅里叶变换)。快速傅里叶变换的公式通常应归功于 Cooley 和 Tukey[1965]。然而,快速傅里叶变换的有趣历史在此值得一提。为响应 Cooley 和 Tukey 的论文,Rudnick[1966]报道了他采用的一种相似技术,运算量同样为 $M\log_2 M$,是基于 Danielson 和 Lanczos[1942]所发表的一种算法。这些作者依次参考了 Runge[1903,1905]的论著作为他们的技术来源。稍后的两篇论文和 Runge 与 König[1924]的讲稿包含了现代快速傅里叶变换算法的基本优点。相似的技术还由 Yates[1937],Stumpff[1939],Good[1958]和 Thomas[1963]发表过。在 1965 年,Cooley-Tukey 论文发表前,一篇 Cooley,Lewis 与 Welch[1967a]合作的论文中进行了历史性的概括和对结果有趣的对比。

习题

★4.1 式(4.2.5)和式(4.2.6)所示的$F(u)$与$f(x)$是一组傅里叶变换对。对于$f(x)$,将式(4.2.6)代入式(4.2.5),会发现左右两边相等。再重复该过程,对于$F(u)$,将式(4.2.5)代入式(4.2.6)。需要用到下面的指数正交性质:

$$\sum_{x=0}^{M-1} e^{j2\pi rx/M} e^{-j2\pi ux/M} = \begin{cases} M & r = u \\ 0 & \text{其他} \end{cases}$$

4.2 证明傅里叶变换和傅里叶反变换是线性过程(见 2.6 节的线性部分)。

4.3 令$F(u,v)$表示一幅图像的 DFT。由 4.2.3 节的讨论可知，$F(u,v)$乘上一个滤波器函数$H(u,v)$，再进行傅里叶反变换将改变图像的外观，这依赖于滤波器的固有特性。假设 $H(u,v)=A$，A 为正常数。滤波的结果将是图像乘以这个常数。应用卷积定理，在数学上解释为什么空间域的像素表达式会乘上同一个常数。

★4.4 高斯型低通滤波器在频率域中的传递函数是：

$$H(u,v) = Ae^{-(u^2+v^2)/2\sigma^2}$$

证明空间域的相应滤波器形式为：

$$h(x,y) = A\sqrt{2\pi}\sigma e^{-2\pi^2\sigma^2(x^2+y^2)}$$

（提示：将变量当做连续的，以简化计算。）

4.5 如式(4.4.1)所示，高通滤波器的传递函数为：

$$H_{hp}(u,v) = 1 - H_{lp}(u,v)$$

这里，$H_{lp}(u,v)$是相应低通滤波器的传递函数。用习题 4.4 的结论解释什么是空间域高斯高通滤波器函数的形式。

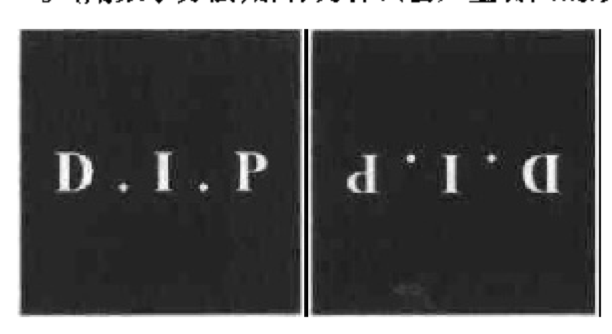
4.6 ★(a)证明式(4.2.21)的正确性。

(b)证明式(4.6.1)和式(4.6.2)的正确性。

4.7 在图 4.11(b)中频率谱的横坐标轴上近似周期性出现的亮点来源是什么？

★4.8 图 4.23 所示的每一个空间域滤波器在原点有一个尖峰，解释其来源。

4.9 观察如下所示的图像。右边的图像这样得到：(a)在原始图像左边乘上$(-1)^{x+y}$；(b)计算离散傅里叶变换(DFT)；(c)对变换取复共轭；(d)计算傅里叶反变换；(e)结果的实部再乘以$(-1)^{x+y}$。(用数学方法)解释为什么会产生右图的效果。



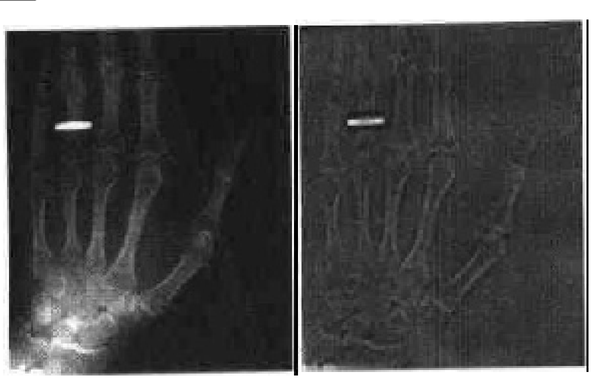
4.10 证明如果一个滤波器传递函数 $H(u,v)$是实函数且对称，那么对应的空间域滤波器$h(x,y)$也是对称实函数。

★4.11 证明卷积定理的正确性。为了简化，只限于单变量的连续函数。

4.12 观察如下所示的图像。右边的图像是对左边图像用高斯型低通滤波器进行低通滤波，然后用高斯型高通滤波器对结果再进行高通滤波得到的。图像的大小为 420 × 344，且两个滤波器的 $D_0=25$。

(a)考虑右图，解释戒指中心部分为什么明亮又是实心的。而过滤后图像的显著特性是，物体(如手指、手腕、骨骼)的外部边沿之间的区域更加黑暗，换句话说，你是不希望高通滤波器勾画戒指暗区内部的恒定区，因为高通滤波排除了直流项吗？

(b)当滤波处理的顺序相反时，你认为结果会有区别吗？



(原图像由密歇根大学医学院解剖学教研室的 Thomas R.Gest 博士提供)

4.13 给出一幅 $M \times N$ 大小的图像,要求做一个实验,它由截频为 D_0 的高斯型低通滤波器重复进行低通滤波。可以忽略计算的舍入误差。令 k_{min} 代表完成实验的机器可表示的最小正数。

★(a) 令 K 代表滤波器使用的次数。在没有实验前,你能预测 K 为足够大的值时结果将是什么吗? 如果能,结果是什么?

(b) 推导出 K 最小值的表达式以保证能够得到预测的结果。

4.14 假设你构成了一个空间域低通滤波器。将点(x,y)直接相邻的 4 个点取平均值,但排除该点本身。

(a) 在频率域中找到等价的滤波器 $H(u,v)$。

(b) 证明你的结果是一个低通滤波器。

★**4.15** 近似一个离散导数(3.7 节)的基本方法包括对 $f(x+1,y)-f(x,y)$取差分。

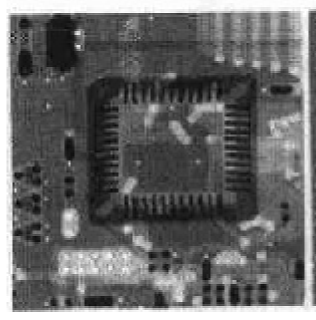
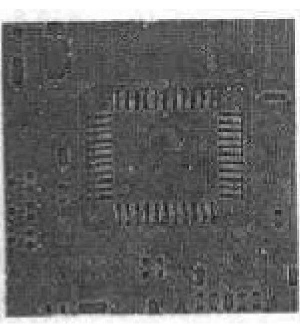
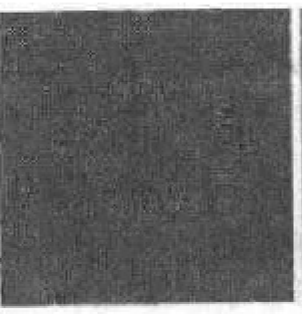
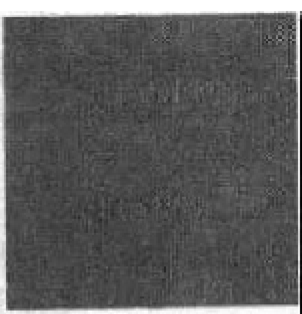
(a) 找到滤波器传递函数 $H(u,v)$,以在频率域中进行等价操作。

(b) 证明 $H(u,v)$是一个高通滤波器。

4.16 考虑下面所示的图像序列。最左边的图是商业印刷电路板的一部分 X 射线图像。右面的图像分别是用一个 $D_0 = 30$ 的高斯高通滤波器进行 1 次、10 次、100 次滤波的结果。图像大小为 330×334 像素点,每个点有 8 比特灰度。为了显示,它们被缩放成适当比例,但对这个问题并没有影响。

(a) 对于这些结果有一个提示,即在有限次过滤后图像将停止改变。说明实际是否是这样。可以忽略舍入误差。令 k_{min} 代表完成此实验的机器可表示的最小正数。

(b) 如果在(a)中确定有限次迭代后变化将停止,求最小的迭代次数。

(原图像由 Lixi 公司的 Joseph E.Pascente 先生提供)

4.17 如图 4.30 所示，将高频加强和直方图均衡相结合是得到边沿锐化和对比度增强的有效方法。

(a) 证明它是否与哪个处理第一次应用有关系。

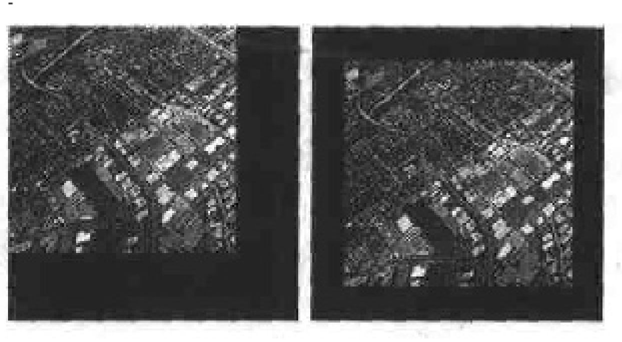
(b) 如果应用顺序有影响，给出一个第一次采用某个方法的理由。

★4.18 能否想出一个用傅里叶变换计算(或部分计算)在图像微分中使用的梯度数值的方法(见 3.7.3 节)？如果你的回答是肯定的，给出一个实现的方法。如果回答是"否"，解释为什么。

4.19 在 4.4.4 节我们以拉普拉斯算子的定义开始，并推导出一个可通过频率域进行等价操作的滤波器。以图 4.27(f)所示的近似模板开始，并在频率域找到实现近似操作的滤波器。

4.20 用一个 n 阶巴特沃思型高通滤波器的传递函数构建一个同态滤波器。这个滤波器必须显示出图 4.32 的特性，并且必须包括图中所示的参数。

★4.21 在 4.6.3 节较详细地讨论了频率域过滤时需要的图像延拓。在该节中，说明了需要延拓的图像在图像中行和列的末尾要填充 0 值(见下面左图)。你认为如果我们把图像放在中心，四周填充 0 值(见下面右图)而不改变 0 值的总数，会有区别吗？请解释。

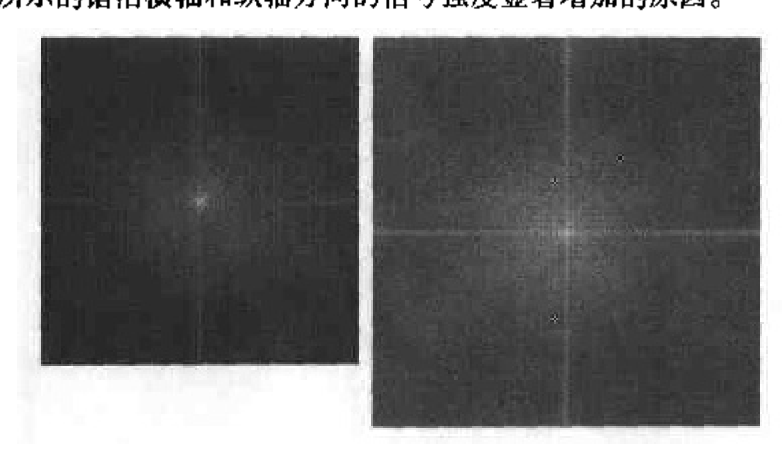


原图像由 NASA 提供

4.22 同一幅图像的两个傅里叶频谱如下图所示。左边的频谱是原始图像的，右边的频谱是图像延拓 0 值后获取的。

(a) 解释总体对比中的不同。

(b) 解释右图所示的谱沿横轴和纵轴方向的信号强度显著增加的原因。



4.23 图 4.41(e)所示的相关函数是根据式(4.6.31)通过频率域计算出来的。图 4.41(d)所示的模板变换的复共轭用于计算。如果用图像变换的共轭来代替，图 4.41(e)将会是什么样子？

4.24 参考图 4.41,如果字母 UTK 只有一个像素偏离以下的各位置,图 4.41(e)所示的相关函数将会是什么样子?

★ (a) 图像的左边界

★ (b) 图像的右边界

(c) 图像的顶部

(d) 图像的底部

在(a)和(b)中,假设字母以通过所有字母中心和图像中心的想像的水平中线为中心。在(c)和(d)中,假设字母以通过字母“T”中心和图像中心的想像的垂直中线为中心。

★**4.25** 证明式(4.6.47)和式(4.6.48)的正确性(提示:用归纳证明法)。

4.26 假设给出一个对恒星事件分析的实验生成的图像集合。每幅图像包含一组明亮且松散的点,对应宇宙广袤空间中的星星。问题是这些星星因为大气折射和阶梯型亮度分布几乎不可见。如果这些图像用一组脉冲模型化成恒定照射分量的乘积,请提出一个增强过程,它基于显示星星自身的图像分量的同态滤波。

4.27 一种成熟的医学技术用于检测电子显微镜生成的某类图像。为了简化检查任务,技术人员决定采用数字图像增强技术检查具有代表性的图像以发现如下问题:(a)不感兴趣的明亮的孤立点;(b)缺乏锐化;(c)一些图像对比度不够;(d)平均灰度值的改变,这时该值应该是 V,以便进行正确的某种强度测量。技术人员要纠正这些问题,在余留的灰度保持正常色调时,用白色显示在 I_1 和 I_2 波段之间的灰度。提出一种处理方案,技术人员可以按该步骤达到目的,可以使用第 3 章、第 4 章的技术。

第5章 图像复原

正如在图像增强中那样，图像复原技术的最终目的是改善给定的图像。尽管图像增强和图像复原有相交叉的领域，但图像增强主要是一个主观的过程，而图像复原的大部分过程是一个客观过程。图像复原试图利用退化现象的某种先验知识来重建或复原被退化的图像。因而，复原技术就是把退化模型化，并且采用相反的过程进行处理，以便复原出原图像。

这些方法通常都会涉及到设立一个最佳的准则，它将会产生期望结果的最佳估计。对比而言，图像增强技术基本上是一个探索性过程，为了人类视觉系统的生理接受特点而设计一种改善图像的方法。例如，增强技术被认为是一种对比度拉伸，因为它主要基于提供给观看者喜欢接受的图像，而通过去模糊函数去除图像模糊则认为是图像复原技术。

本章中介绍的材料是严谨的引导。仅仅是从给出一幅退化数字图像的角度来考虑复原问题，因此，在传感器，数字转换器和显示退化方面考虑的话题比较浅显。尽管这些主题在图像复原应用的整体处理中是非常重要的，但它们超出了本章所要讨论的范围。

正如第3章，第4章所讲的，一些复原技术在空间域已被很好地公式化了，而另一些较好地适用于频域。例如，当退化仅仅是加性噪声时，空间处理就非常适用。而另一方面，如图像模糊一类的退化通过空间域用较小掩模的方法就很困难。在这种情况下，基于不同的最优性准则的频域过滤是可选择的方法。这些滤波器同样也考虑了现存的噪声。如第4章中所讲的那样（见4.6.7的论述），在频域中用来解决给定应用的复原滤波器通常作为生成一个数字滤波器的基础。这些数字滤波器更适合于用硬件或固件来实现。

5.1 图像退化/复原过程的模型

如图5.1所示，在本章中，退化过程中可以被模型化为一个退化函数和一个加性噪声项，处理一幅输入图像$f(x,y)$产生一幅退化图像$g(x,y)$。给定$g(x,y)$和关于退化函数H的一些知识以及外加噪声项$\eta(x,y)$，图像复原的目的是获得关于原始图像的近似估计$\hat{f}(x,y)$。通常我们希望这一估计尽可能接近原始输入图像，并且H和η的信息知道得越多，所得到的$\hat{f}(x,y)$就会越接近$f(x,y)$。在本章中使用的大部分方法是基于不同的图像复原滤波器。

如在5.5节中所示的，如果系统H是一个线性、位置不变性的过程，那么在空间域中给出的退化图像可由下式给出：

$$g(x,y) = h(x,y) * f(x,y) + \eta(x,y) \tag{5.1.1}$$

其中，$h(x,y)$是退化函数的空间描述；与第4章相同，“*”表示空间卷积。从4.2.4节和4.6.4节知道，空间域上的卷积等同于频域上的乘积，因此，可以把式(5.1.1)的模型写成等价的频域下的描述：

$$G(u,v) = H(u,v)F(u,v) + N(u,v) \tag{5.1.2}$$

公式中大写字母项是式(5.1.1)中相应项的傅里叶变换。这两个公式是本章中大部分材料的

基础。

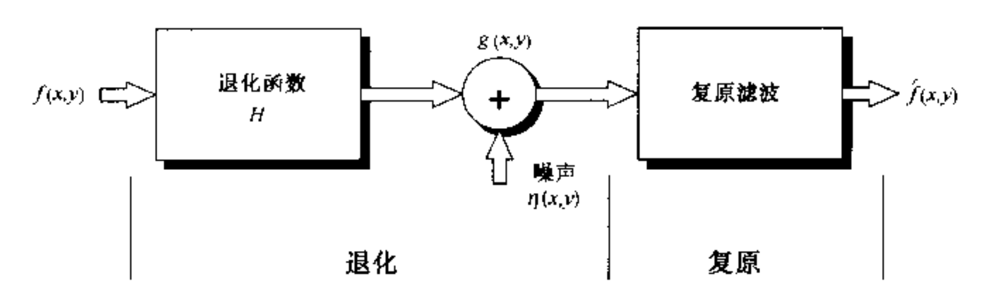


图 5.1 图像退化/复原过程的模型

在后面,我们假设 H 是一个等同的运算符,并且只处理由噪声引起的退化。从 5.6 节开始,考虑若干重要的图像退化函数,而且考虑几个 H 和 η 都存在的情况下的图像复原方法。

5.2 噪声模型

数字图像的噪声主要来源于图像的获取(数字化过程)和传输过程。图像传感器的工作情况受各种因素的影响,如图像获取中的环境条件和传感元器件自身的质量。例如,使用 CCD 摄像机获取图像,光照程度和传感器温度是生成图像中产生大量噪声的主要因素。图像在传输过程中主要由于所用的传输信道的干扰受到噪声污染。比如,通过无线网络传输的图像可能会因为光或其他大气因素的干扰被污染。

5.2.1 噪声的空间和频率特性

相关的讨论是定义噪声空间特性的参数和这些噪声是否与图像相关。频率特性是指噪声在傅里叶域的频率内容（即，相对于电磁波谱），例如，当噪声的傅里叶谱是常量时，噪声通常称为白噪声。这个术语是从白光的物理特性派生出来的，它将以相等的比例包含可见光谱中所有的频率。从第 4 章的讨论中不难看出，以等比例包含所有频率的函数的傅里叶谱是一个常量。

由于空间的周期噪声的异常(5.2.3 节),在本章中假设噪声独立于空间坐标,并且它与图像本身无关联(简言之,噪声分量值和像素值之间不相关)。这些假设至少在某些应用中(有限量子成像,例如 X 光和核医学成像就是一个很好的例子)是无效的,但复杂的处理空间非独立和相关噪声的情况不在我们所讨论的范围。

5.2.2 一些重要噪声的概率密度函数

基于前面章节的假设,所关心的空间噪声描述符是 5.1 节中所提及模型的噪声分量灰度值的统计特性。它们可以被认为是由概率密度函数(PDF)表示的随机变量,下面是在图像处理应用中最常见的 PDF。

高斯噪声

由于高斯噪声在空间和频域中数学上的易处理性,这种噪声(也称为正态噪声)模型经常被用于实践中。事实上,这种易处理性非常方便,使高斯模型经常用于临界情况下。

高斯随机变量 z 的 PDF 由下式给出：

$$p(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(z-\mu)^2/2\sigma^2} \tag{5.2.1}$$

其中，z 表示灰度值，μ 表示 z 的平均值或期望值，σ 表示 z 的标准差。标准差的平方 σ^2，称为 z 的方差. 高斯函数的曲线如图 5.2(a)所示。当 z 服从式(5.2.1)的分布时，其值有 70% 落在 $[(\mu-\sigma),(\mu+\sigma)]$ 范围内，且有 95% 落在 $[(\mu-2\sigma),(\mu+2\sigma)]$ 范围内。

瑞利噪声

瑞利噪声的概率密度函数由下式给出：

$$p(z) = \begin{cases} \frac{2}{b}(z-a)e^{-(z-a)^2/b} & z \geq a \\ 0 & z < a \end{cases} \tag{5.2.2}$$

概率密度的均值和方差由下式给出：

$$\mu = a + \sqrt{\pi b/4} \tag{5.2.3}$$

和

$$\sigma^2 = \frac{b(4-\pi)}{4} \tag{5.2.4}$$

图 5.2(b)显示了瑞利密度的曲线。注意，距原点的位移和其密度图形的基本形状向右变形的事实。瑞利密度对于近似偏移的直方图十分适用。

伽马(爱尔兰)噪声

伽马噪声的 PDF 由下式给出：

$$p(z) = \begin{cases} \frac{a^b z^{b-1}}{(b-1)!}e^{-az} & z \geq 0 \\ 0 & z < 0 \end{cases} \tag{5.2.5}$$

其中，$a>0$，b 为正整数且“!”表示阶乘。其密度的均值和方差由下式给出：

$$\mu = \frac{b}{a} \tag{5.2.6}$$

和

$$\sigma^2 = \frac{b}{a^2} \tag{5.2.7}$$

图 5.2(c)显示了伽马密度的曲线，尽管式(5.2.5)经常被用来表示伽马密度，严格地说，只有当分母为伽马函数 $\Gamma(b)$ 时才是正确的。当分母如表达式所示时，该密度近似称为爱尔兰密度。

指数分布噪声

指数噪声的 PDF 可由下式给出：

$$p(z) = \begin{cases} ae^{-az} & z \geq 0 \\ 0 & z < 0 \end{cases} \tag{5.2.8}$$

其中，$a>0$。概率密度函数的期望值和方差是：

$$\mu = \frac{1}{a} \tag{5.2.9}$$

$$\sigma^2 = \frac{1}{a^2} \tag{5.2.10}$$

注意,指数分布的概率密度函数是当 $b = 1$ 时爱尔兰概率分布的特殊情况。图 5.2(d)显示了该密度函数的曲线。

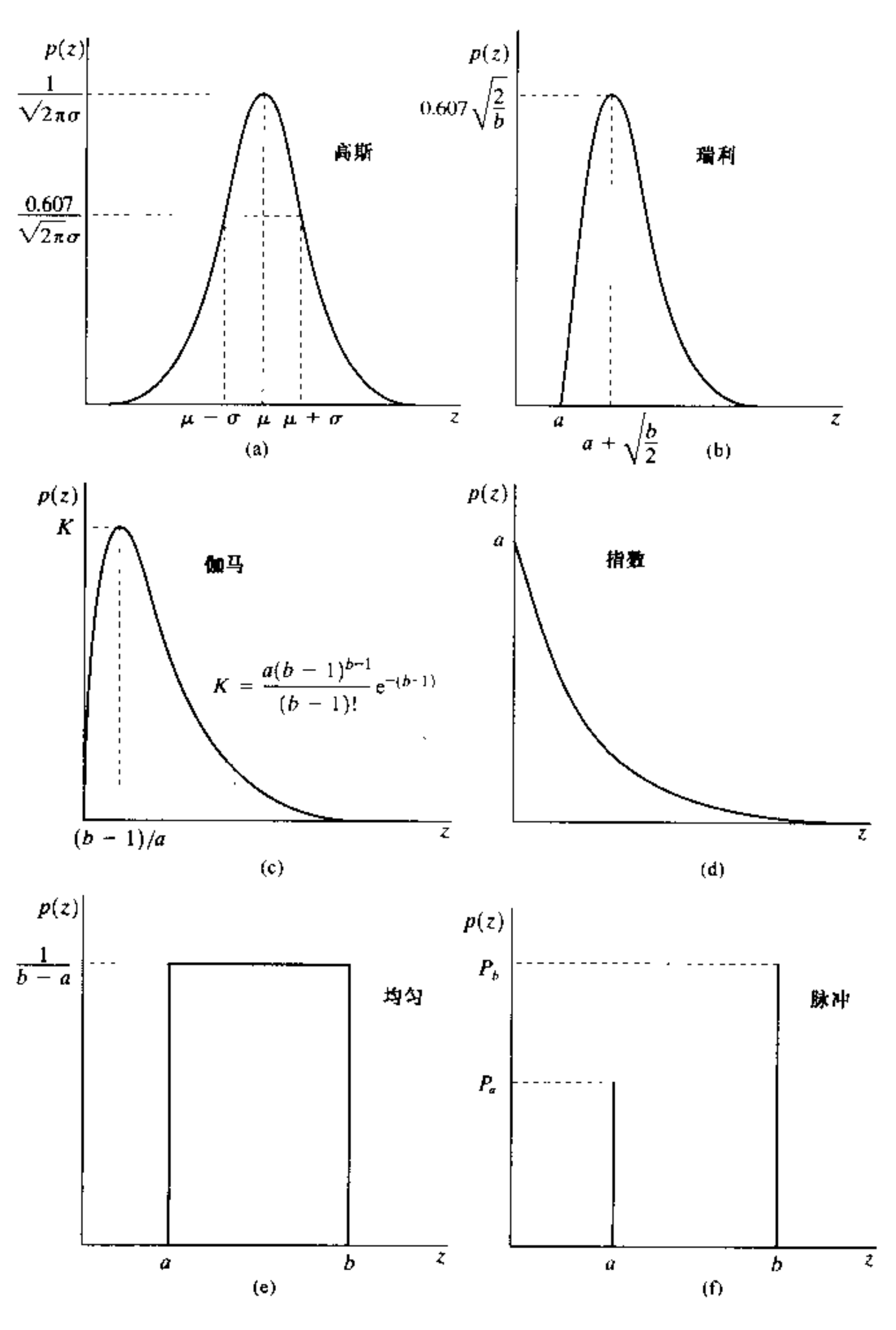


图 5.2 某些重要的概率密度函数

均匀分布噪声

均匀分布噪声的概率密度,可由下式给出:

$$p(z) = \begin{cases} \frac{1}{b-a} & a \leqslant z \leqslant b \\ 0 & \text{其他} \end{cases} \tag{5.2.11}$$

概率密度函数的期望值和方差可由下式给出:

$$\mu = \frac{a+b}{2} \tag{5.2.12}$$

$$\sigma^2 = \frac{(b-a)^2}{12} \tag{5.2.13}$$

图5.2(e)显示了均匀密度的曲线。

脉冲噪声(椒盐噪声)

(双极)脉冲噪声的PDF可由下式给出:

$$p(z) = \begin{cases} P_a & z = a \\ P_b & z = b \\ 0 & \text{其他} \end{cases} \tag{5.2.14}$$

如果 $b > a$,灰度值 b 在图像中将显示为一个亮点,相反,a 的值将显示为一个暗点。若 P_a 或 P_b 为零,则脉冲噪声称为单极脉冲。如果 P_a 和 P_b 均不可能为零,尤其是它们近似相等时,脉冲噪声值将类似于随机分布在图像上的胡椒和盐粉微粒。由于这个原因,双极脉冲噪声也称为椒盐噪声。同时,它们有时也称为散粒和尖峰噪声。在我们的讨论中,将交替使用脉冲噪声和椒盐噪声这两个术语。

噪声脉冲可以是正的,也可以是负的。标定通常是图像数字化过程的一部分。因为脉冲干扰通常与图像信号的强度相比较大,因此,在一幅图像中,脉冲噪声总是数字化为最大值(纯黑或纯白)。这样,通常假设 a, b 是饱和值,从某种意义上看,在数字化图像中,它们等于所允许的最大值和最小值。由于这一结果,负脉冲以一个黑点(胡椒点)出现在图像中。由于相同的原因,正脉冲以白点(盐点)出现在图像中。对于一个8位图像,这意味着 $a = 0$(黑),$b = 255$(白)。图5.2(f)显示了脉冲噪声的概率密度函数。

前述的一组PDF为在实践中模型化宽带噪声干扰状态提供了有用的工具。例如,在一幅图像中,高斯噪声的产生源于电子电路噪声和由低照明度或高温带来的传感器噪声。瑞利密度分布在图像范围内特征化噪声现象时非常有用。指数密度分布和伽马密度分布在激光成像中有一些应用。像前几章所提及的那样,脉冲噪声主要表现在成像中的短暂停留中,例如,错误的开关操作。均匀密度分布可能是在实践中描述得最少的,然而,均匀密度作为模拟随机数产生器的基础是非常有用的。

例5.1 样本噪声图像和它们的直方图

图5.3显示了一个非常适合于阐述刚刚所讨论的噪声模型的测试图。之所以选择它,是因为它是由简单、恒定的区域所组成,且其从黑到近似于白仅仅有3个灰度级增长跨度。

这方便了对附加在图像上的各种噪声分量特性的视觉分析。

图 5.4 显示了叠加了本节讨论的 6 种噪声的测试图。所示的每幅图像的下面是从图像直接计算而来的直方图。在每种情况下选择噪声的参数,这样对应于测试图中 3 种灰度的直方图将开始合并。这可以使噪声十分显著,而不会使构成图像的基本结构变暗。

比较图 5.4 的直方图和图 5.2 中的概率密度函数,可以看到相近的对应关系。椒盐噪声实例的直方图在光谱的白端有一个额外的尖峰,因为噪声分量是纯黑或纯白,并且在测试图中最亮的分量(圆)是亮灰度。除了少许亮度不同外,在图 5.4 中很难区别出前 5 幅图像有什么显著的不同,即使它们的直方图有明显的区别。椒盐噪声是惟一一种引起退化的视觉可见的噪声类型。

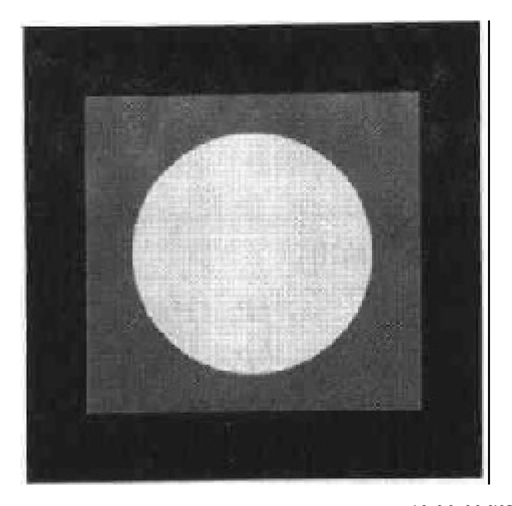


图 5.3 用于说明示于图 5.2 的噪声 PDF 特性的测试图

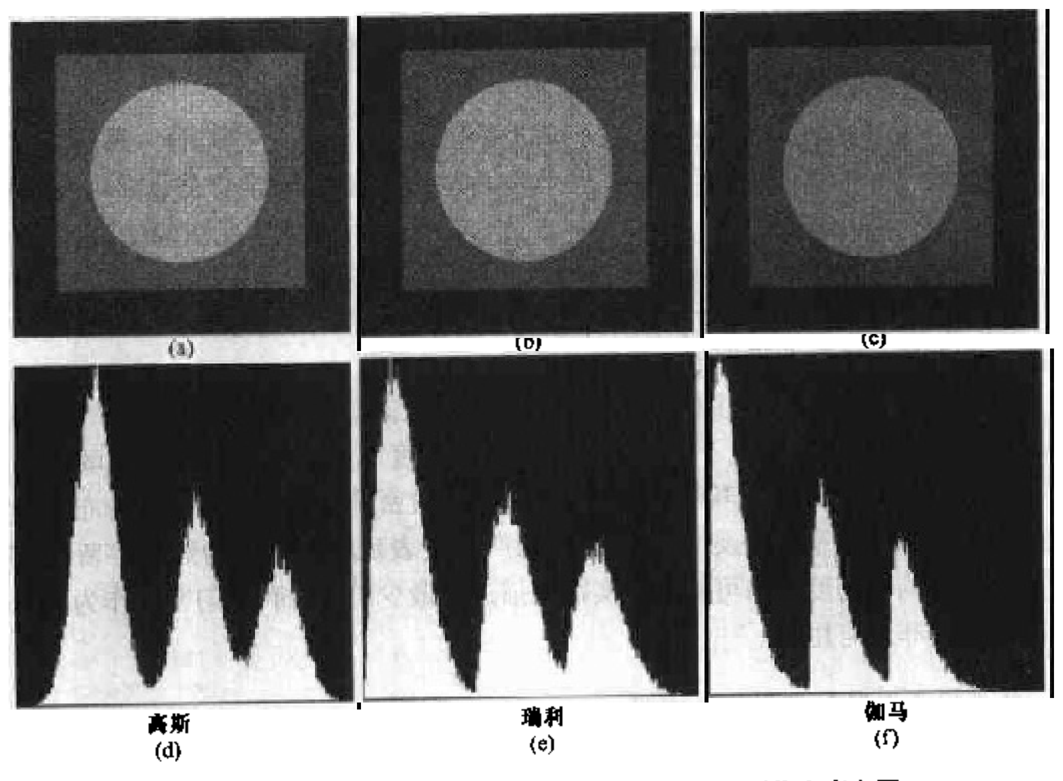


图 5.4 在图 5.3 中附加高斯、瑞利和伽马噪声后的图像和直方图

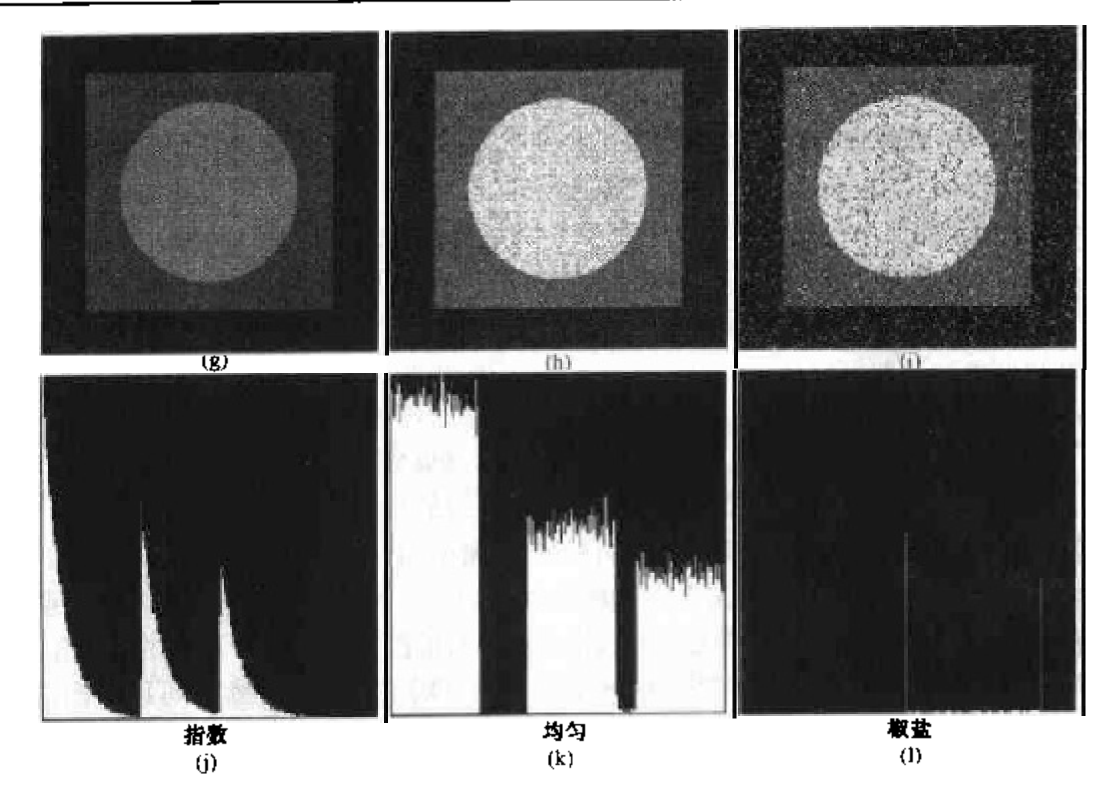


图 5.4(续)　在图 5.3 中附加指数、均匀和椒盐噪声的图像和直方图

5.2.3 周期噪声

在一幅图像中,周期噪声是在图像获取中从电力或机电干扰中产生的。这是惟一的一种空间依赖型噪声,将在本章中讨论。如同 5.4 节中讨论的那样,周期噪声通过频域滤波可以显著地减少。例如,考虑图 5.5(a)所示的图像。这一图像被不同频率的正弦噪声严重干扰了。一个纯正弦的傅里叶变换是位于正弦波共轭频率处的一对共轭脉冲(如表 4.1)。因此,如果在空间域上,正弦波的振幅足够强,将在图像谱中看到对应图像中每个正弦波的脉冲对。如图 5.5(b)所示,确实如此,由于在这个特殊例子中频率值是这样安排的,脉冲以近似于圆的形状出现。将在 5.4 节进一步讨论此问题和关于周期噪声的其他例子。

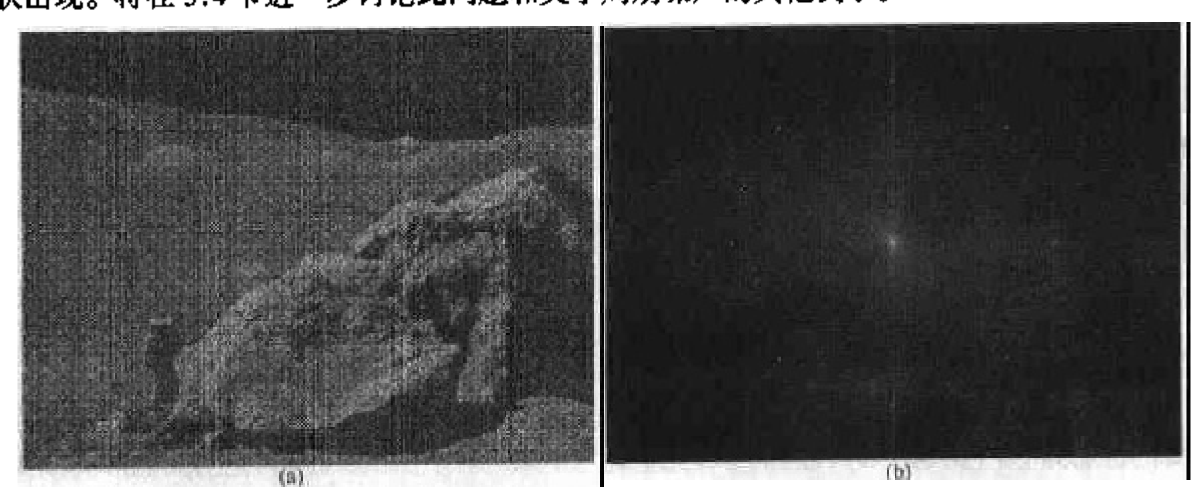


图 5.5　(a)由正弦噪声污染的图像,(b)谱(与一个正弦波相对应的每一对共轭脉冲)(原图像由 NASA 提供)

5.2.4　噪声参数的估计

典型的周期噪声参数是通过检测图像的傅里叶谱来进行估计的。像在前几节提及的那样,周期噪声趋向于产生频率尖峰,这些尖峰甚至通过视觉分析也经常可以检测到。另一种方法是尽可能直接从图像中推断噪声分量的周期性,但这仅仅在非常简单的情况下才是可能的。当噪声尖峰格外显著或可以使用关于干扰的频率分量一般位置的某些知识时,自动分析是可能的。

噪声 PDF 参数一般可以从传感器的技术说明中得知,但对于特殊的成像装置常常有必要去估计这些参数。如果成像系统可用,那么研究这个系统的噪声特性最简单的方法就是截取一组“平坦”环境的图像。例如,在光学传感器情况下,这就像对一个固体的、光照均匀的灰度板成像一样简单。结果图像是一个典型的系统噪声良好的指示器。

当仅仅通过传感器产生的图像可以利用的时候,常常可以从合理的恒定灰度值的一小部分估计 PDF 的参数。例如,在图 5.6 中所示的垂直带(150×20 像素)是从图 5.4 中高斯、瑞利和均匀图像中获取的。所显示的直方图是通过这些小带的图像数据计算出来的。与图 5.6 中的直方图相对应的图 5.4 中的直方图是图 5.4(d),(e),(k)三组中的一组。可以看出,这些相应的直方图形状非常接近于图 5.6 中的直方图形状。由于缩放,它们的高度不同,但它们的形状明显类似。

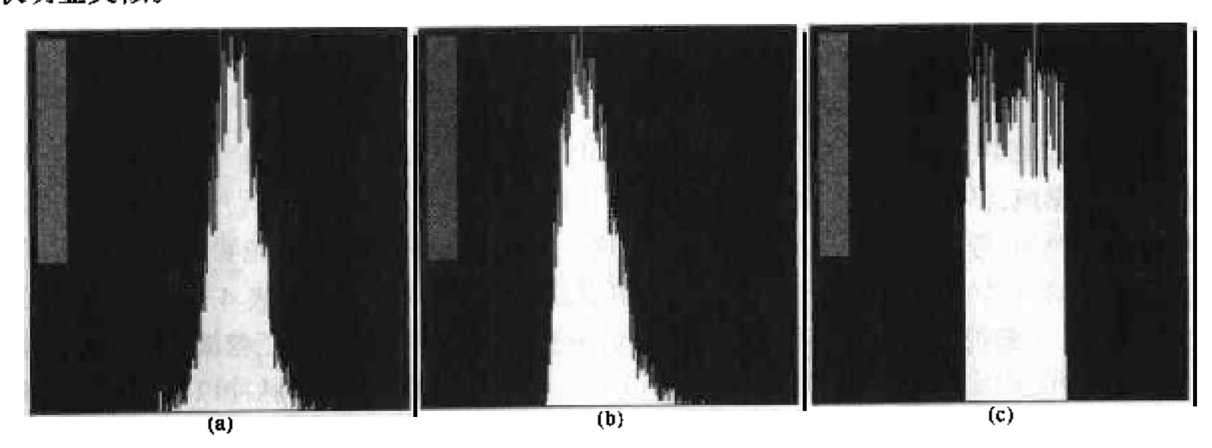


图 5.6　用小带(如图中插入的小图所示)从图 5.4 中的(a)高斯,(b)瑞利,(c)均匀噪声计算的直方图

利用图像带中的数据最简单的方法是计算灰度值的均值和方差。考虑由 S 定义的一条小带(子图像)。可以从基本统计量出发利用下面的样本近似:

$$\mu = \sum_{z_i \in S} z_i p(z_i) \tag{5.2.15}$$

$$\sigma^2 = \sum_{z_i \in S} (z_i - \mu)^2 p(z_i) \tag{5.2.16}$$

其中 z_i 值是 S 中像素的灰度值,且 $P(z_i)$表示相应的归一化直方图值。

直方图的形状指出最接近的 PDF 匹配。如果其形状近似于高斯,那么均值和方差正是所需要的,因为高斯 PDF 可以通过两个参数完全确定下来。对于在 5.2.2 节讨论的其他形状,用均值和方差来解出参数 a 和 b。脉冲噪声用不同的方法处理,因为需要估计黑、白像素发生的实际概率。获得这些估计值需要黑白像素是可见的,因此,为了计算直方图,图像中一个相对恒定的

中等灰度区域是必需的。对应于黑、白像素的尖峰高度是在式(5.2.14)中的 P_a 和 P_b 的估计值。

5.3 噪声存在下的惟一空间滤波复原

当在一幅图像中惟一存在的退化是噪声时，式(5.1.1)和式(5.1.2)变成：

$$g(x,y) = f(x,y) + \eta(x,y) \tag{5.3.1}$$

和

$$G(u,v) = F(u,v) + N(u,v) \tag{5.3.2}$$

噪声项是未知的，从 $g(x,y)$或 $G(u,v)$中减去它们不是一个现实的选择。在周期噪声的情况下，通常从 $G(u,v)$的谱去估计 $N(u,v)$是可能的，如在 5.2.3 节所讲的那样。这时，从 $G(u,v)$中减去 $N(u,v)$得到原始图像的估计。然而，通常这种知识仅仅属于例外而不是规律。

当仅有加性噪声存在时，可以选择空间滤波方法。将结合 3.5 节和 3.6 节中的图像增强做详细的说明。事实上，在这一特殊情况下，图像的增强和复原几乎是不可区别的，除了通过一种特殊的滤波来计算特性以外，执行所有滤波的机理完全如在 3.5 节中讨论过的那样。

5.3.1 均值滤波器

在本节中，简要讨论在 3.6 节中介绍的用于减少噪声的空间滤波器并研究一些其他滤波器，这些滤波器的性能在许多情况下优于 3.6 节所讨论的滤波器。

算术均值滤波器

这是最简单的均值滤波器。令 S_{xy}表示中心在(x,y)点，尺寸为 $m\times n$ 的矩形子图像窗口的坐标组。算术均值滤波过程就是计算由 S_{xy}定义的区域中被干扰图像$g(x,y)$的平均值。在任意点(x,y)处复原图像 $\hat{f}$ 的值就是用S_{xy}定义的区域的像素计算出来的算术均值。即：

$$\hat{f}(x,y) = \frac{1}{mn}\sum_{(s,t)\in S_{xy}} g(s,t) \tag{5.3.3}$$

这个操作可以用其系数为 $1/mn$ 的卷积模板来实现。就像在 3.6.1 节所讨论的那样，均值简单地平滑了一幅图像的局部变化。在模糊了结果的同时减少了噪声。

几何均值滤波器

用几何均值滤波器复原的一幅图像由如下表达式给出：

$$\hat{f}(x,y) = \left[\prod_{(s,t)\in S_{xy}} g(s,t)\right]^{\frac{1}{mn}} \tag{5.3.4}$$

其中，每一个被复原像素由子图像窗口中像素点的乘积并自乘到 $1/mn$ 次幂给出。如例 5.2 中所示，几何均值滤波器所达到的平滑度可以与算术均值滤波器相比，但在滤波过程中会丢失更少的图像细节。

谐波均值滤波器

使用谐波均值滤波器的操作由如下表达式给出：

$$\hat{f}(x,y) = \frac{mn}{\sum_{(s,t)\in S_{xy}} \frac{1}{g(s,t)}} \tag{5.3.5}$$

谐波均值滤波器对于“盐”噪声效果更好,但是不适用于“胡椒”噪声。它善于处理像高斯噪声那样的其他噪声。

逆谐波均值滤波器

逆谐波均值滤波操作对一幅图像的复原基于表达式:

$$\hat{f}(x,y)=\frac{\sum_{(s,t)\in S_{xy}} g(s,t)^{Q+1}}{\sum_{(s,t)\in S_{xy}} g(s,t)^{Q}} \tag{5.3.6}$$

其中 Q 称为滤波器的阶数。这种滤波器适合减少或是在实际中消除椒盐噪声的影响。当 Q 值为正数时,滤波器用于消除“胡椒”噪声;当 Q 值为负数时,滤波器用于消除“盐”噪声。但它不能同时消除这两种噪声。注意,当 $Q=0$ 时,逆谐波均值滤波器退变为算术均值滤波器;当 $Q=-1$ 时,逆谐波均值滤波器退变为谐波均值滤波器。

例 5.2 均值滤波的说明

图 5.7(a)显示了一块电路板的 X 射线图像,图 5.7(b)显示了相同的图像,但被附加的均值为 0、方差为 400 的高斯噪声污染了。对于这种类型的图像,这是非常严重的噪声。图 5.7(c)和图 5.7(d)分别显示了经过 3×3 算术均值滤波器和同样尺寸的几何均值滤波器滤除噪声的结果。尽管这两种噪声滤波器对噪声的衰减都起到了作用,但几何均值滤波器并没有像算术均值滤波器那样使图像变得模糊。例如,图像顶端的连接片在图 5.7(d)中比在图 5.7(c)中更为清晰。对于图像的其他部分也是如此(见习题 5.1)。

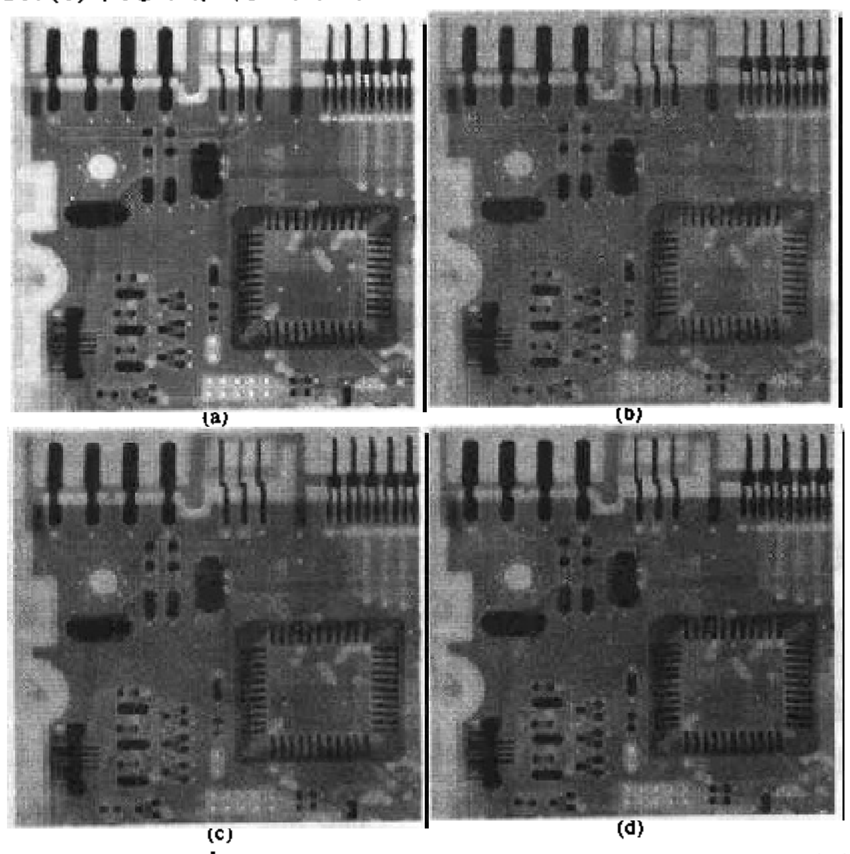


图 5.7 (a)X 射线图像,(b)由附加高斯噪声污染的图像,(c)用 3×3 算术均值滤波器滤波的结果,(d)用相同尺寸的几何均值滤波器滤波的结果(原图像由Lixi公司的Joseph E. Pascente先生提供)

图 5.8(a)显示了同样的电路图像,但有 0.1 的概率被"胡椒"噪声干扰。类似地,图 5.8(b)显示了同样的电路图,但有 0.1 的概率被"盐"噪声干扰。图 5.8(c)显示了图 5.8(a)经 $Q=1.5$ 的逆谐波均值滤波器处理后的结果。图 5.8(d)显示了图5.8(b)经 $Q=-1.5$ 的逆谐波均值滤波器处理后的结果。两种滤波器都有很好的去除噪声效果,这种正阶滤波器在使暗区模糊的损失下,使背景较为清晰,对于负阶滤波器,情况则刚好相反。

总的来说,算术均值滤波器和几何均值滤波器(尤其是后者)更适合于处理高斯或均匀等随机噪声。谐波均值滤波器更适合于处理脉冲噪声,但它还有一个缺点,就是必须知道噪声是暗噪声还是亮噪声,以便于选择合适的 Q 符号。如果 Q 的符号选择错了可能会引起灾难性后果,如图 5.9 所示。接下来讨论的一些滤波器可以消除此缺点。

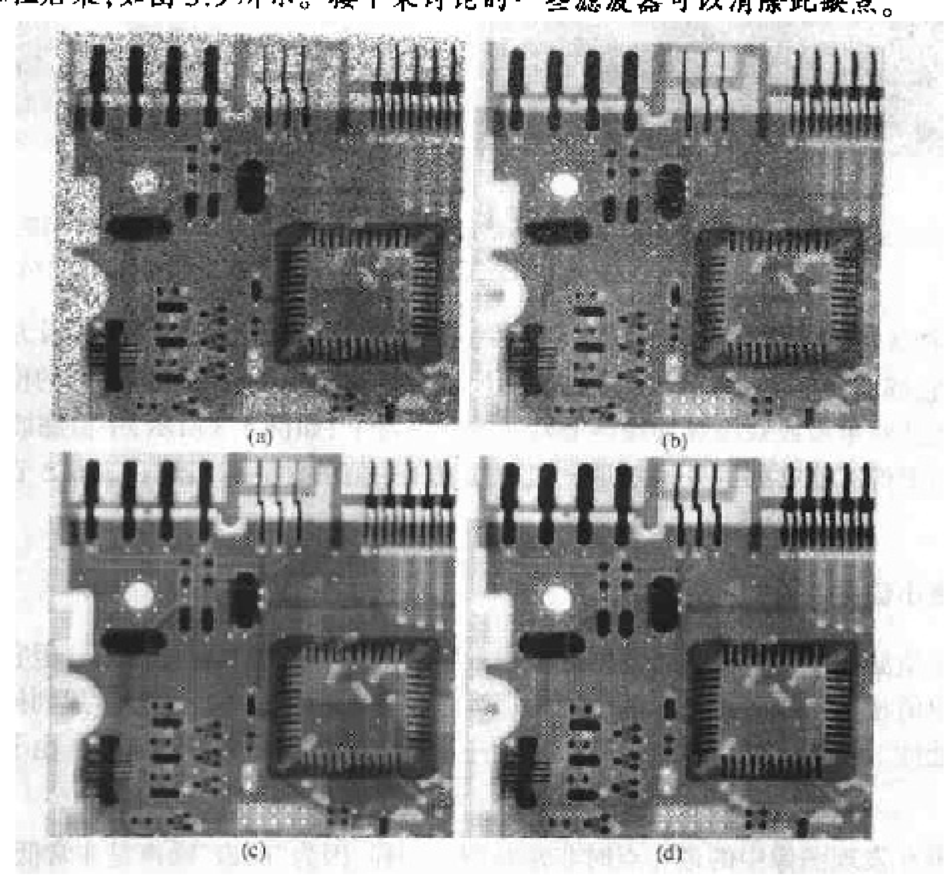


图 5.8　(a)以 0.1 的概率被"胡椒"噪声污染了的图像,(b)以相同概率被"盐"噪声污染的图像,(c)用3×3大小、阶数为1.5的逆谐波滤波器滤波的结果,(d)用 $Q=-1.5$滤波(b)的结果

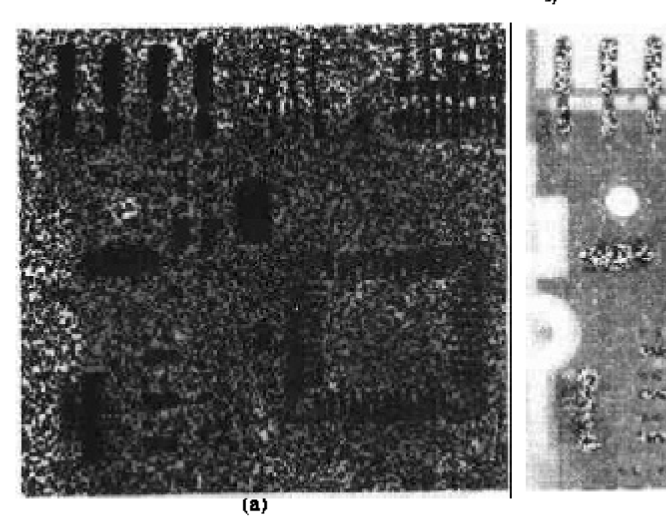
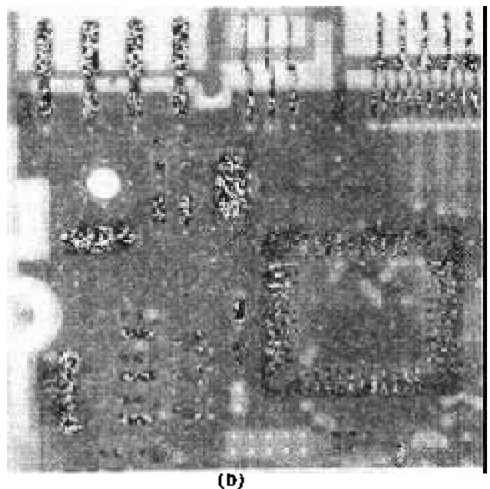
5.3.2 顺序统计滤波器

3.6.2 节介绍过了顺序统计滤波器。在本节中对其进行扩充并介绍一些其他的顺序统计滤波器。如在 3.6.2 节中所讲的那样,顺序统计滤波器是空间域滤波器,它们的响应基于由滤波器包围的图像区域中像素点的排序。滤波器在任何点的响应由排序结果决定。

中值滤波器

最著名的顺序统计滤波器是中值滤波器,就像其名字那样,用该像素的相邻像素的灰度中值来替代该像素的值:

$$\hat{f}(x,y) = \underset{(s,t)\in S_{xy}}{\text{median}}\{g(s,t)\} \tag{5.3.7}$$

(a) (b)

图5.9 在逆谐波滤波中错误选择符号的结果。(a)用3×3大小和 $Q=-1.5$ 的逆谐波滤波器对图5.8(a)滤波的结果,(b)用 $Q=1.5$ 滤波图5.8(b)的结果

像素的原始值包含在中值的计算结果中,中值滤波器的应用非常普遍,因为对于很多种随机噪声,它都有良好的去噪能力,且在相同尺寸下比起线性平滑滤波器引起的模糊较少。中值滤波器尤其对单极或双极脉冲噪声非常有效。实际上,如例5.3所示,中值滤波器对于这种噪声干扰的图像有非常好的处理效果。这种滤波器中值的计算和实现在3.6.2节中有详细的讨论。

最大值和最小值滤波器

尽管中值滤波器是目前为止图像处理中最常用的一种顺序统计滤波器,但它绝不是惟一的一种。中值相当于顺序排列数值中间的那个数,读者会从基本的统计学想到排列还有很多其他的可能性。例如,可以使用序列中最后一个数值,得出最大值滤波器,由下式给出:

$$\hat{f}(x,y) = \max_{(s,t)\in S_{xy}}\{g(s,t)\} \tag{5.3.8}$$

这种滤波器在发现图像中的最亮点时非常有用。同样,因为"胡椒"噪声是非常低的值,作为子图像区域 S_{xy} 的最大值选择结果,它可以通过这种滤波器消除。

起始位置的滤波器称为最小值滤波器:

$$\hat{f}(x,y) = \min_{(s,t)\in S_{xy}}\{g(s,t)\} \tag{5.3.9}$$

这种滤波器对发现图像中的最暗点非常有用。同样,作为最小值操作的结果,它可以用来消除"盐"噪声。

中点滤波器

中点滤波器是在滤波器涉及范围内计算最大值和最小值之间的中点:

$$\hat{f}(x,y) = \frac{1}{2}\left[\max_{(s,t)\in S_{xy}}\{g(s,t)\} + \min_{(s,t)\in S_{xy}}\{g(s,t)\}\right] \tag{5.3.10}$$

注意,这种滤波器结合了顺序统计和求平均,对于高斯和均匀随机分布这类噪声有最好的效果。

修正后的阿尔法均值滤波器

假设在 S_{xy} 邻域内去掉 $g(s,t)$ 最高灰度值的 $d/2$ 和最低灰度值的 $d/2$。用 $g_r(s,t)$ 来代表剩余的 $mn-d$ 个像素。由这些剩余像素点的平均值形成的滤波器称为修正后的阿尔法均值滤波器。

$$\hat{f}(x,y)=\frac{1}{mn-d}\sum_{(s,t)\in S_{xy}}g_r(s,t) \tag{5.3.11}$$

其中，d 值可以取 0 到 $mn-1$ 之间的任意数。当 $d=0$ 时，修正后的阿尔法均值滤波器退变为前面几节中所讨论的算术均值滤波器。如果选择 $d=(mn-1)/2$，修正后的阿尔法均值滤波器将退变为中值滤波器。d 取其他值时，修正后的阿尔法均值滤波器在包括多种噪声的情况下非常适用，例如高斯噪声和椒盐噪声混合的情况下。

例 5.3 顺序统计滤波器的说明

图 5.10(a)显示了被概率为 $P_a=P_b=0.1$ 的脉冲噪声干扰的电路图像。图 5.10(b)显示了用规格为 3×3 的中值滤波器处理后的结果。与图 5.10(a)所示图像相比较，改进是显而易见的，但是一些噪声点仍然是可见的。第二次通过(对图 5.10(b)的图像)中值滤波器处理后去掉了大部分这样的噪声点，仅剩余非常少的可见噪声点。这些噪声点在经过第三次中值滤波器处理后全部消除。这是一个很好的说明处理脉冲型加性噪声的中值滤波器能力的例子。请记住，重复地使用中值滤波器处理可能会使图像模糊化，所以应尽可能保持所希望的处理次数。

图 5.11(a)显示了利用最大值滤波器对图 5.8(a)图像中的"胡椒"噪声进行处理后的图像。这种滤波器对除去图像中的"胡椒"噪声的确很合适，但也注意到它同时也从黑色物体的边缘移走了一些黑色像素。图 5.11(b)显示了利用最小值滤波器处理图5.8(b)图像的结果。在这种特殊情况下，对于噪声消除，最小值滤波器的确比最大值滤波器更好，但它同时也从亮色物体的边缘移走了一些白色像素。这样就使亮色物体变小，而同时使暗色物体变大(就像在图像顶端的连接片)，这是因为围绕这些物体的白点被设置成了暗灰度级。

接下来阐述修正后的阿尔法均值滤波器，图 5.12(a)显示了一幅被均值为 0，方差为 800 的加性均匀噪声干扰的图像。这是高噪声级别的污染，它被 $P_a=P_b=0.1$ 的椒盐噪声叠加后进一步劣化了，如图 5.12(b)所示。在图像中，对于这种强噪声干扰，认为通过更大的滤波器来处理是正确的。图 5.12(c)到图 5.12(f)显示了依次用算术均值、几何均值、中值和 $d=5$ 且规格为 5×5 的修正后的阿尔法均值滤波器处理后的图像。如所预料的一样，由于脉冲噪声的存在，算术均值滤波器和几何均值滤波器(尤其是后者)并没有起到良好的作用。中值滤波器和修正后的阿尔法均值滤波器则好得多，在噪声消除方面，阿尔法均值滤波器做得更好些。例如，图像左上角第 4 个连接片在阿尔法均值滤波器处理后变得稍微平滑一些。这并不意外，因为对于大的 d 值，阿尔法均值滤波器的性能接近于中值滤波器，并仍然存在一些平滑能力。

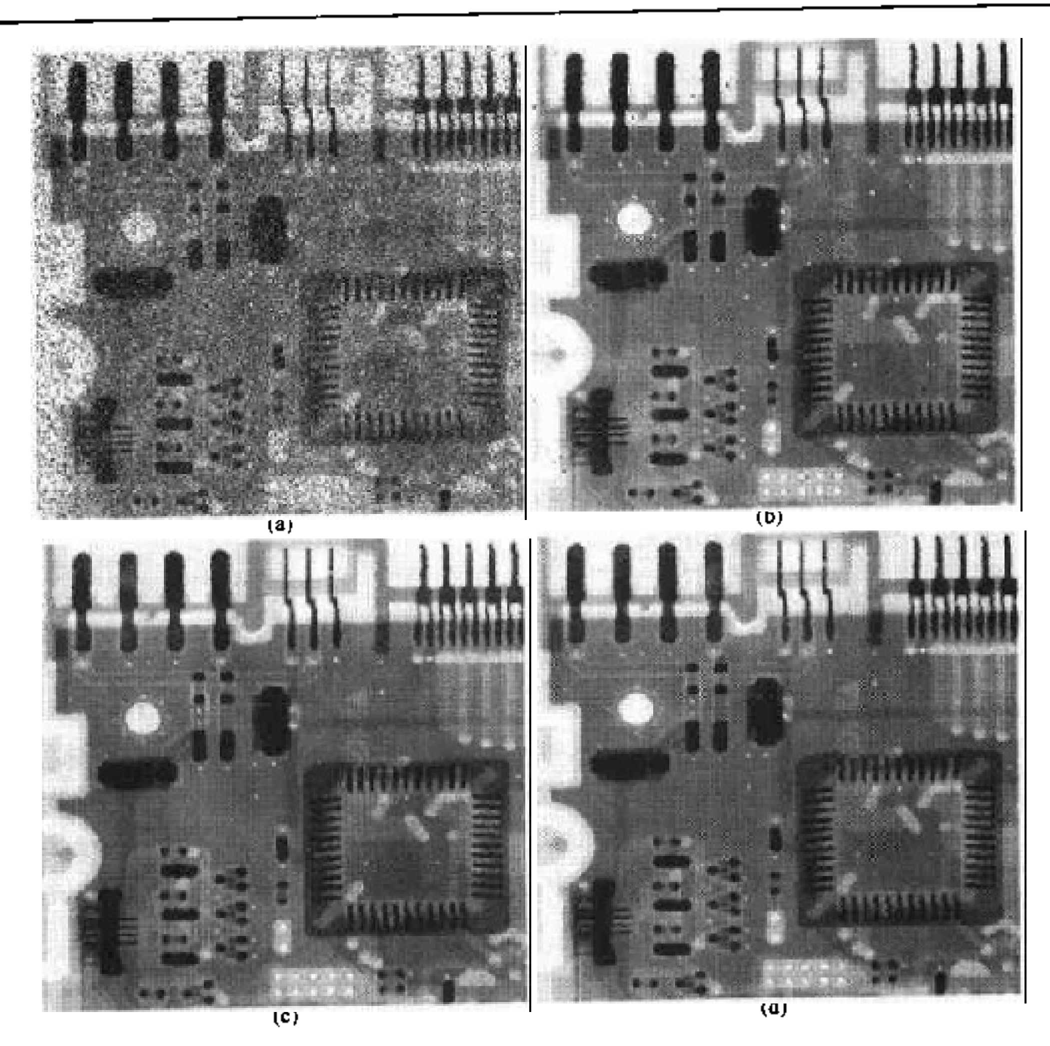


图 5.10 (a)由概率为 $P_a = P_b = 0.1$ 的椒盐噪声污染的图像,(b)用尺寸为 3×3 的中值滤波器处理的结果,(c)用该滤波器处理(b)的结果,(d)用相同的滤波器处理(c)的结果

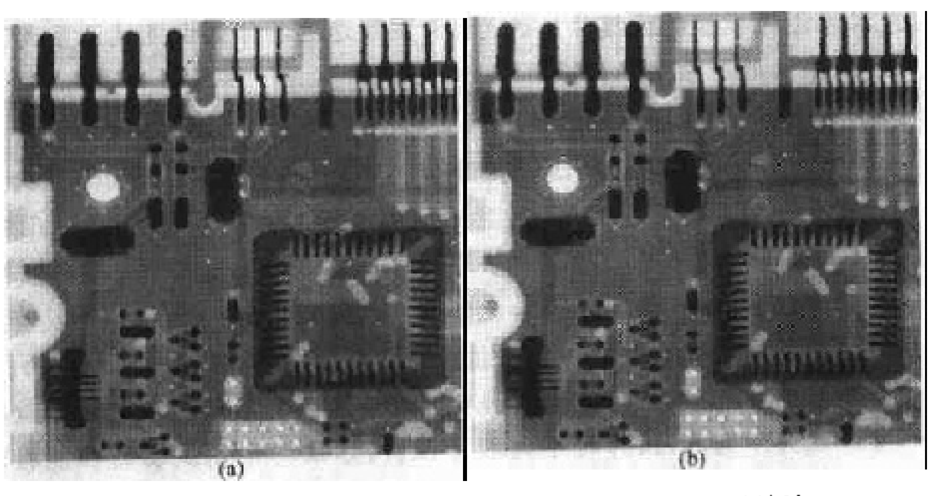


图 5.11 (a)用大小为 3×3 的最大值滤波器对图 5.8(a)滤波的结果,(b)用最小值滤波器对图 5.8(b)滤波的结果

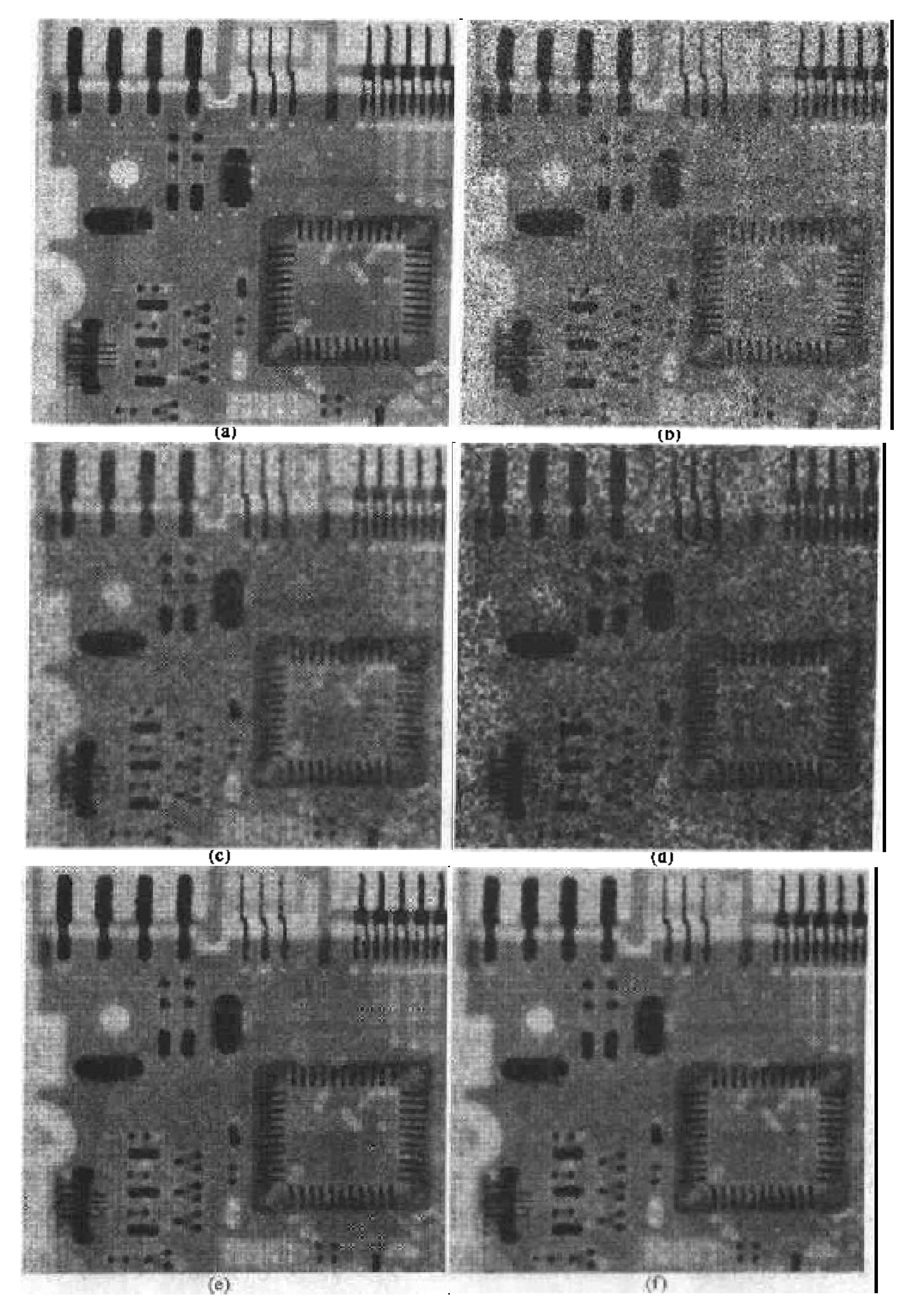


图 5.12 (a)由加性均匀噪声污染的图像,(b)由加性椒盐噪声污染的图像。图(b)中的图像用如下的 5×5 滤波器滤波:(c)算术均值滤波器,(d)几何均值滤波器,(e)中值滤波器,(f) $d=5$ 的修正后的阿尔法均值滤波器

5.3.3 自适应滤波器

迄今为止讨论过的滤波器被选择应用于图像中后,并没有考虑图像中的一点对于其他点

的特征有什么不同。在这一节中,将看一下两个简单的自适应滤波器,它们的行为变化基于由 $m \times n$ 矩形窗口 S_{xy} 定义的区域内图像的统计特性。正如下边的讨论所显示的那样,自适应滤波器要优于迄今为止讨论过的所有滤波器的性能。作为提高滤波能力的代价是滤波器的复杂度。请记住,我们一直在处理这样的情况,即退化图像等于原始图像加噪声,而并没有考虑其他的退化类型。

自适应、局部噪声消除滤波器

随机变量最简单的统计度量是均值和方差。这些适当的参数是自适应滤波器的基础,因为它们是与图像状态紧密相关的数据。均值给出了计算均值的区域中灰度平均值的度量,而方差给出了这个区域的平均对比度的度量。

滤波器作用于局部区域 S_{xy}。滤波器在中心化区域中任何点 (x,y) 上的滤波器响应基于以下 4 个量:

(a) $g(x,y)$ 表示噪声图像在点 (x,y) 上的值;

(b) σ_η^2,干扰 $f(x,y)$ 以形成 $g(x,y)$ 的噪声方差;

(c) m_L,在 S_{xy} 上像素点的局部均值;

(d) σ_L^2,在 S_{xy} 上像素点的局部方差。

滤波器的预期性能如下:

1. 如果 σ_η^2 为零,滤波器应该简单地返回 $g(x,y)$ 的值。这无关紧要,在 $g(x,y)$ 下零噪声的情况等同于 $f(x,y)$。

2. 如果局部方差与 σ_η^2 是高相关的,那么滤波器要返回一个 $g(x,y)$ 的近似值。一个典型的高局部方差是与边缘相关的,并且这些边缘应该保留。

3. 如果两个方差相等,希望滤波器返回区域 S_{xy} 上像素的算术均值。这种情况发生在局部面积与全部图像有相同特性的条件下,并且局部噪声简单地用求平均来降低。

为了获得 $\hat{f}(x,y)$,基于这些假定的自适应表达式可以写成:

$$\hat{f}(x,y) = g(x,y) - \frac{\sigma_\eta^2}{\sigma_L^2}[g(x,y) - m_L] \tag{5.3.12}$$

惟一需要知道成估计的量是全部噪声的方差,σ_η^2。其他参数需要从 S_{xy} 中各个坐标 (x,y) 处的像素计算出来,在该位置滤波器窗口已被中心化。在式(5.3.12)中,假设 $\sigma_\eta^2 \leqslant \sigma_L^2$。模型中的噪声是加性和位置独立的,因此,这是一个合理的假设,因为 S_{xy} 是 $g(x,y)$ 的子集。然而,我们很少有确切的 σ_η^2 的知识,因此,在实际中很可能违反这个条件。由于这个原因,对式(5.3.12)的实现应构建一个测试,以便如果条件 $\sigma_\eta^2 > \sigma_L^2$ 发生,把比率设置为 1。这使该滤波器为非线性的。然而,它可以防止由于缺乏图像噪声方差的知识而产生的无意义结果(即,负灰度值,依赖于 m_L 值)。另一种方法是允许产生负值,并在最后重新标定灰度值。这个结果将损失图像的动态范围。

例 5.4 自适应局部噪声消除滤波的说明

图 5.13(a)显示了一幅电路图像,被均值为 0 且方差为 1 000 的加性高斯噪声所污染。这是严重的噪声污染,但是它提供了一个理想的测试平台,在这个平台上可以比较相关滤波

器的性能。图 5.13(b)是使用大小为 7×7 的算术均值滤波器处理噪声图像的结果。噪声被平滑掉了，但是代价是图像严重地模糊了。在图 5.13(c)中也同样如此，图 5.13(c)显示了使用几何均值滤波器处理噪声图像的结果，该滤波器大小也为 7×7。这两幅滤波图像间的不同之处与例 5.2 所讨论的相似；只是模糊的程度不同。

图 5.13 (d)显示了使用式(5.3.12)中 $\sigma_\eta^2=1\,000$ 的自适应滤波器的结果。这一结果的改进与前面两个滤波器相比是很大的。从所有的噪声消减情况来看，自适应滤波器所达到的效果与算术和几何均值滤波器相似。然而，用自适应滤波器滤波的图像更尖锐些。例如，在图5.13(d)中图像顶部的连接片更尖锐一些。而其他特性，如孔、图像左下方暗部件的八条腿等特征在图5.13 (d)中更加清楚。这些效果在自适应滤波器处理的图像中是非常典型的。正如在本章开始提到过的，改进性能所付出的代价是增加滤波器的复杂性。

以上结果中对于 σ_η^2 用了一个值，该值准确地与噪声方差匹配。若该值未知，且使用了太低的估计值，则算法会因为其校正量比应有的小而返回非常接近原图的图像。估计值太高，会造成方差比在 1.0 处中断，并且算法会比正常时更频繁地从图像中减去平均值。如果允许为负值，且图像在最后被重新标定，则结果就会如前所述，将会损失动态范围。

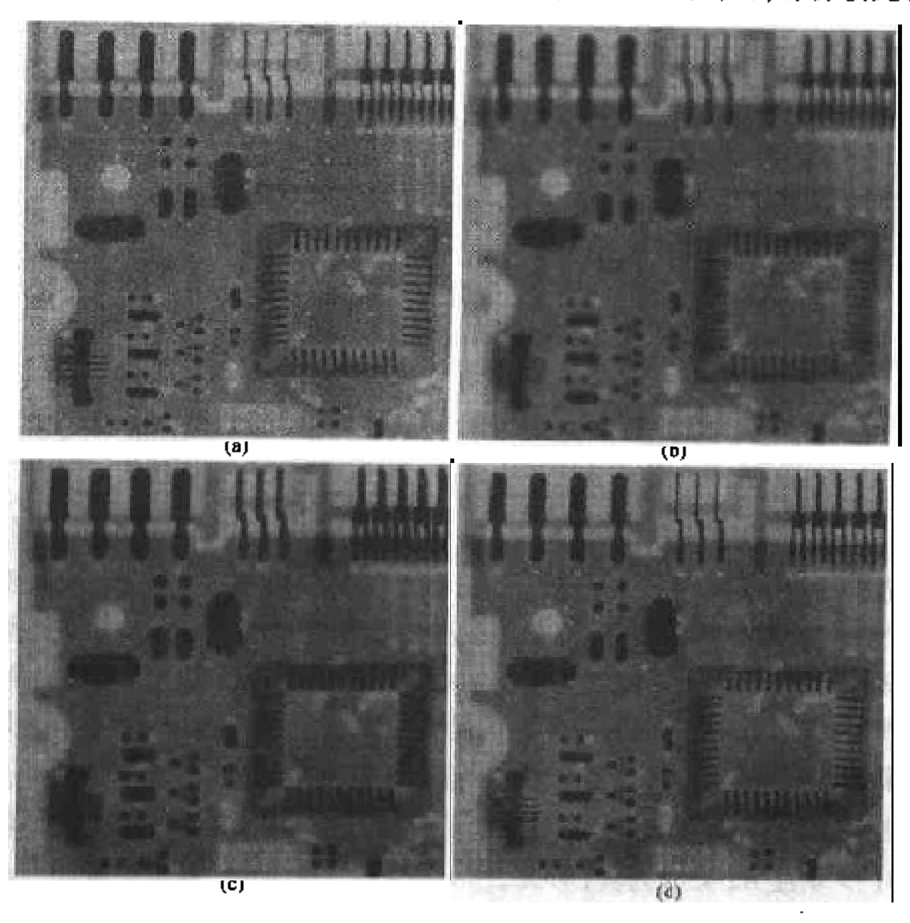


图 5.13 (a)由零均值和方差为 1 000 的加性高斯噪声污染的图像，(b)算术均值滤波的效果，(c)几何均值滤波效果，(d)自适应噪声削减滤波的效果。所有滤波器的大小为7×7

自适应中值滤波器

5.3.2 节所讨论的中值滤波器只要冲激噪声的空间密度不大,性能将会很好(根据经验,P_a,P_b小于 0.2)。本节中显示自适应中值滤波器可以处理具有更大概率的冲激噪声。自适应中值滤波器的另一个优点是,平滑非冲激噪声时可以保存细节,这是传统中值滤波器做不到的。正如前面章节讨论的所有滤波器那样,自适应中值滤波器也能工作在矩形窗口区 S_{xy}上。然而,不像这些滤波器,自适应中值滤波器当进行滤波处理时,依赖本节列举的一定条件而改变(或提高)S_{xy}的大小。请记住,滤波器的输出是一个单值,该值用于代替点(x,y)处的像素值,点(x,y)是在给定时间窗口 S_{xy}被中心化后的一特殊点。

考虑如下符号:

z_{min} = S_{xy}中灰度级的最小值

z_{max} = S_{xy}中灰度级的最大值

z_{med} = S_{xy}中灰度级的中值

z_{xy} = 在坐标(x,y)上的灰度级

S_{max} = S_{xy}允许的最大尺寸

自适应中值滤波器算法工作在两个层次,定义为 A 层和 B 层,如下所示:

A 层: $A1 = z_{med} - z_{min}$

$A2 = z_{med} - z_{max}$

如果 $A1 > 0$ 且 $A2 < 0$,转到 B 层

否则增大窗口尺寸

如果窗口尺寸$\leqslant S_{max}$重复 A 层

否则输出 z_{xy}

B 层: $B1 = z_{xy} - z_{min}$

$B2 = z_{xy} - z_{max}$

如果 $B1 > 0$ 且 $B2 < 0$,输出 z_{xy}

否则输出 z_{med}

理解该算法机理的关键在于,要记住它有 3 个主要目的:除去“椒盐”噪声(冲激噪声),平滑其他非冲激噪声,并减少诸如物体边界细化或粗化等失真。z_{min}和 z_{max}的值进行统计后被算法认为是类冲激式的噪声成分,即使它们在图像中并不是最低和最高的可能像素值。

利用这些观察结果,我们看到,A 层的目的是决定中值滤波器的输出 z_{med}是否是一个脉冲(黑或白)。如果条件 $z_{min} < z_{med} < z_{max}$有效,根据前节提到的原因 z_{med}就不是脉冲。在该情况下,转到 B 层,检测,看一下窗口 z_{xy}中心点本身是否是一个脉冲(回顾一下,z_{xy}是将被处理的点)。若条件 $B1 > 0$ 且 $B2 < 0$ 为真,那么 $z_{min} < z_{xy} < z_{max}$,$z_{xy}$和$z_{med}$就不是脉冲,原因与 z_{med}不是脉冲相同。在这种情况下,算法输出一个不变的像素值 z_{xy}。通过不改变这些“中间水平”的点,来减少图像中的失真。如果条件 $B1 > 0$ 且 $B2 < 0$ 为假,则 $z_{xy} = z_{min}$或 $z_{xy} = z_{max}$。在任一种情况下,像素值都是一个极端值且算法输出中值 z_{med},从 A 层知道 z_{med}不是噪声脉冲。最后一步是标准的中值滤波器所做的。问题是,标准中值滤波器用图像中相应邻域的中值代替该

图像中的每一点,会引起不必要的细节损失。

继续上面的说明，假设 A 层确实找到了一个脉冲（即，如不满足，则测试转到 B 层）。然后，算法会增大窗口尺寸并重复 A 层。该环路会继续直到算法找到一个非脉冲（转到 B 层）中值，或者达到窗口最大尺寸。如果达到了最大窗口尺寸，算法返回 z_{xy} 值。注意，不能保证该值不是一个脉冲。噪声的概率 P_a 和/或 P_b 越小，或者 S_{max} 在允许的范围内越大，过早退出条件发生的可能性就会越小。这似乎是合理的。随着脉冲的密度增大，需要更大的窗口消除尖峰噪声。

算法每输出一个值，窗口 S_{xy} 就被移到图像的下一个位置。然后算法重新初始化，在新的像素位置应用。如习题 3.20 所指出的，使用新像素的中值可以反复更新，因面减少了计算总开销。

例 5.5　自适应中值滤波器的说明

图 5.14(a)显示了被"椒盐"噪声污染了的电路图像,该噪声的概率为 $P_a = P_b = 0.25$,是图 5.10(a)所用噪声的 2.5 倍。在这里,噪声水平非常高,能够模糊图像的大部分细节。作为比较的基础,图像首先使用最小中值滤波器进行滤波,消除大部分可见的脉冲噪声痕迹。需要一个 7×7 的中值滤波器去做这件事,并且图 5.14(b)显示了其效果。虽然噪声被有效消除了,但是滤波器在图像上也引起了明显的细节损失。例如,图像顶部的一些连接片失真或改变了。其他的图像细节也同样被歪曲了。

图 5.14(c)显示了使用 $S_{max} = 7$ 的自适应中值滤波器的效果。噪声消除水平同中值滤波器相似。但是,自适应滤波器保持了点的尖锐性和其细节,从图 5.14(b)看出,其改进是很明显的。连接片没有多少失真,在中值滤波器中被模糊或失真得无法辨认的特征,在图 5.14(c)中变得更尖锐并被更清楚地显示。两个明显的例子就是穿过主板的填充小白孔,以及图像左下四分之一处有八条腿的黑色部件。

考虑到图 5.14(a)的高水平噪声,自适应算法的性能已经非常好了。允许的最大窗口大小的选择与应用有关,但是合理的起始值可以首先通过使用各种尺寸的标准中值滤波器的实验估计获得。这将根据自适应算法性能的期望值建立可视的原始资料。

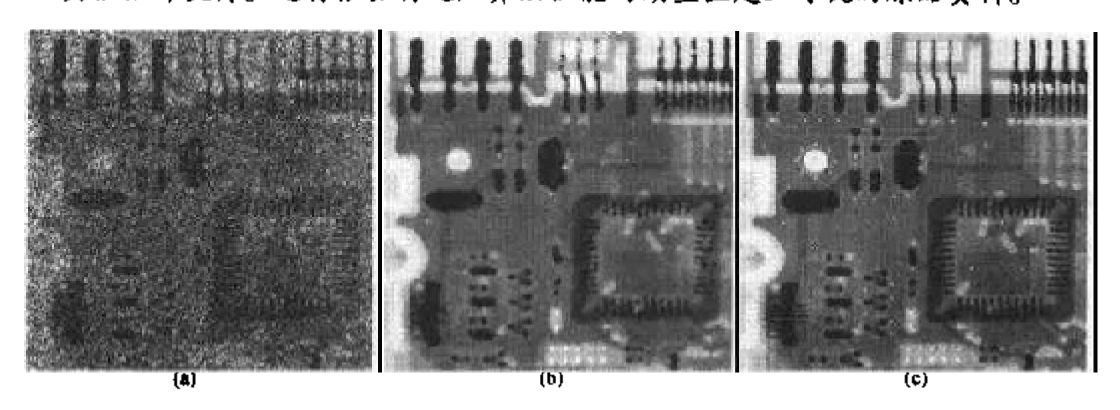


图 5.14　(a)被概率 $P_a = P_b = 0.25$ 的椒盐噪声污染了的图像,(b) 7×7 中值滤波器的滤波效果,(c) $S_{max} = 7$ 的自适应中值滤波器的效果

5.4 频域滤波削减周期噪声

在第4章讨论了低通和高通频域滤波器,把它们作为图像增强的基本工具。本节将讨论更加专用的带阻、带通和陷波滤波器,它们能削减或消除周期性噪声。

5.4.1 带阻滤波器

带阻滤波器消除或衰减了傅里叶变换原点处的频段。理想带阻滤波器的表达式为:

$$H(u,v)=\begin{cases}1, & D(u,v)<D_0-\dfrac{W}{2}\\ 0, & D_0-\dfrac{W}{2}\leqslant D(u,v)\leqslant D_0+\dfrac{W}{2}\\ 1, & D(u,v)>D_0+\dfrac{W}{2}\end{cases} \tag{5.4.1}$$

这里,$D(u,v)$与式(4.3.3)给出的一样,是到中心化频率矩形原点的距离,W是频带的宽度,D_0是频带的中心半径。

同样,n阶的巴特沃思带阻滤波器的表达式为:

$$H(u,v)=\frac{1}{1+\left[\dfrac{D(u,v)W}{D^2(u,v)-D_0^2}\right]^{2n}} \tag{5.4.2}$$

高斯带阻滤波器的表达式为:

$$H(u,v)=1-\mathrm{e}^{-\frac{1}{2}\left[\frac{D^2(u,v)-D_0^2}{D(u,v)W}\right]^2} \tag{5.4.3}$$

图5.15显示了这3个滤波器的透视图。

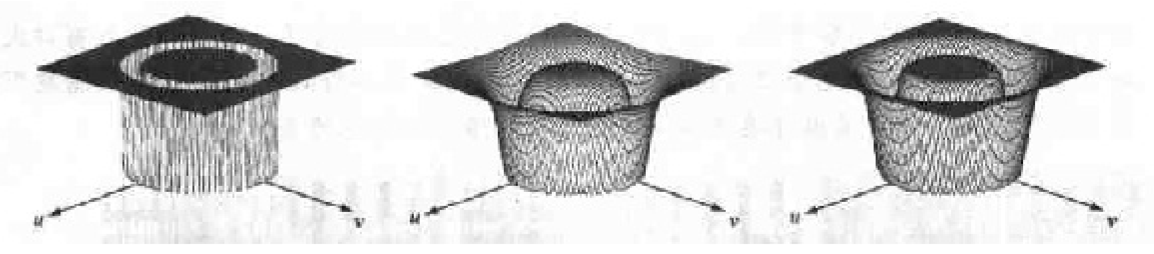


图5.15 从左到右,依次为理想滤波器,巴特沃思滤波器(阶数为1)和高斯带阻滤波器的透视图

例5.6 利用带阻滤波器消除周期性噪声

带阻滤波器的主要应用之一是,在频率域噪声分量的一般位置近似已知的应用中消除噪声。一个典型的例子就是一幅被加性周期性噪声污染的图像,该噪声可被近似为二维正弦函数。不难看出(习题5.14),纯正弦曲线的傅里叶变换是一对脉冲。一个脉冲位于正弦曲线的频率上,而另一个在那个点的镜像图上。两个脉冲实际上都是虚的(正弦曲线的傅里叶变换的实部为0),而且彼此为复共轭。在5.4.3节和5.4.4节还将详述该问题,目前先讲解带阻滤波器。

图5.16(a)与图5.5(a)相同,显示了被不同频率的正弦噪声严重污染的图像。噪声成分很容易被看成图5.16(b)中显示的傅里叶频谱中对称的亮点对。本例中,噪声分量位于关

于变换原点的近似圆上，因此，使用圆对称带阻滤波器是一个正确的选择。图5.16(c)显示了4阶的巴特沃思带阻滤波器，它设置了适当的半径和宽度，完全包围了噪声脉冲。由于通常希望从变换中尽可能小地削减细节，因此在带阻滤波器中，通常要求尖锐、窄的滤波器。使用该滤波器滤波图5.16(a)的效果显示在图5.16(d)中。其改进是非常明显的。即使细小的细节和纹理也被这一简单的滤波方式有效修复了。还应注意到使用小卷积模板的直接空间域滤波方式也不可能取得相同的效果。

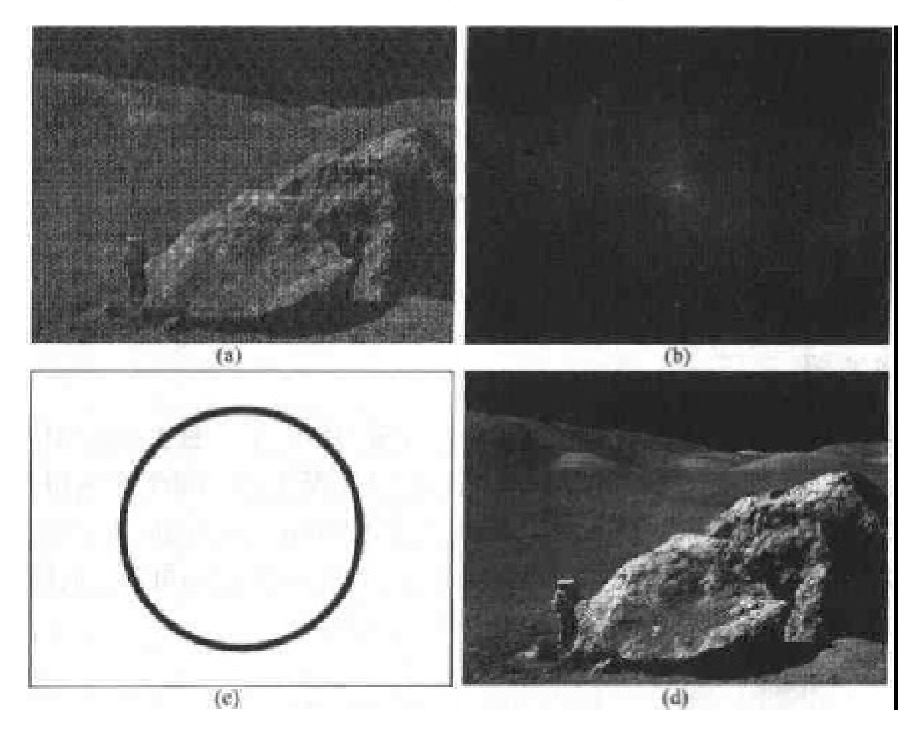


图5.16 (a)被正弦噪声污染的图像，(b)图(a)的频谱，(c)巴特沃思带阻滤波器(白色代表1)，(d)滤波效果图[原图像由NASA提供]

5.4.2 带通滤波器

带通滤波器执行与带阻滤波器相反的操作。在4.4节已经显示了如何使用式(4.4.1)从相应的低通滤波器改为高通滤波器。同样，带通滤波器的传递函数 $H_{bp}(u,v)$ 是根据相应的带阻滤波器的传递函数 $H_{br}(u,v)$ 并应用下式：

$$H_{bp}(u,v) = 1 - H_{br}(u,v) \tag{5.4.4}$$

得到的。把它留给读者作为练习(习题5.12)，根据式(5.4.1)到式(5.4.3)推导带通滤波器的表达式。

例5.7 使用带通滤波器提取噪声模式

在一幅图像上直接执行带通滤波器不是一个通常的做法，因为，这通常会消除太多图像细节。不过，带通滤波在选中频段的图像中屏蔽效果时是非常有用的，这是图5.17中的说明。产生图像的方法有：(1)应用式(5.4.4)获得与前例中使用的带阻滤波器对应的带通滤波器，(2)做带通滤波变换的反向转换。使用上面的方法会丢失图像的大部分细节，不过保留的信息是很有用的。很清楚，使用这种方法复原的噪声模式与污染图5.16(a)的噪

声非常接近。换句话说,带通滤波器帮助屏蔽噪声模式。这是一个有用的结果,因为它简化了噪声分析,相当于与图像内容无关。

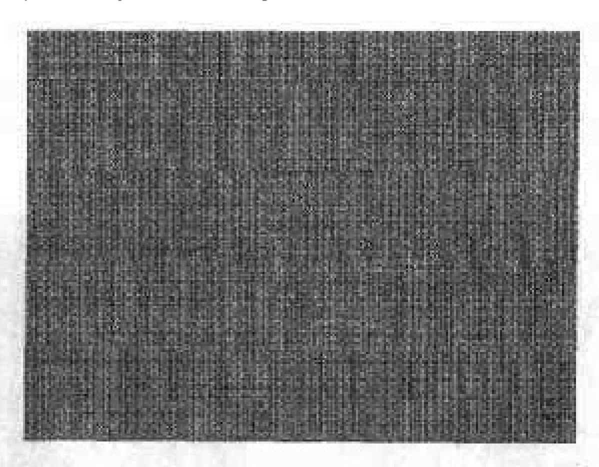


图 5.17 由带通滤波获得的图 5.16(a)图像的噪声模式

5.4.3 陷波滤波器

陷波滤波器阻止(或通过)事先定义的中心频率邻域内的频率。图 5.18 分别显示了理想、巴特沃思和高斯陷波(带阻)滤波器的三维图。由于傅里叶变换是对称的,要获得有效结果,陷波滤波器必须以关于原点对称的形式出现。这个原则的特例是,如果陷波滤波器位于原点处,则要以它本身的形式出现。虽然为说明起见,我们只列举了一对,但是可实现的陷波滤波器的对数是任意的。陷波区域的形状也是任意的(例如,矩形)。

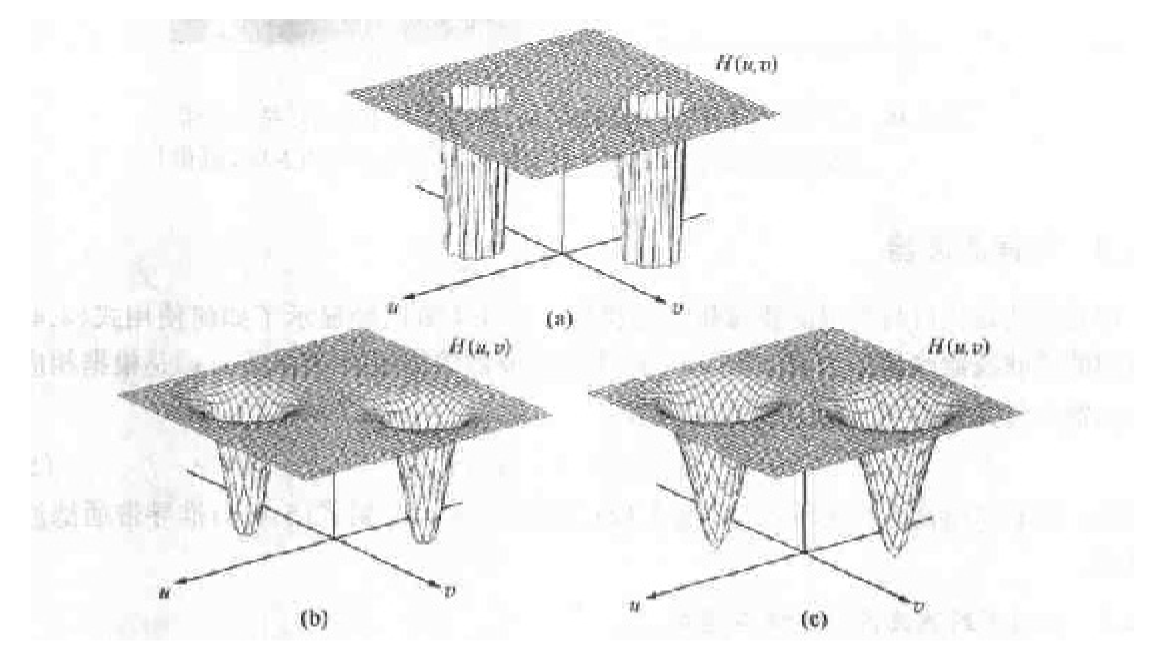


图 5.18 (a)理想的,(b)巴特沃思(2 阶)和(c)高斯陷波(带阻)滤波器的透视图

半径为 D_0,中心在(u_0, v_0)且在$(-u_0, -v_0)$对称的理想陷波带阻滤波器的传递函数为:

$$H(u,v) = \begin{cases} 0 & D_1(u,v) \leqslant D_0 \text{或} D_2(u,v) \leqslant D_0 \\ 1 & \text{其他} \end{cases} \tag{5.4.5}$$

这里，

$$D_1(u,v) = [(u - M/2 - u_0)^2 + (v - N/2 - v_0)^2]^{1/2} \tag{5.4.6}$$

$$D_2(u,v) = [(u - M/2 + u_0)^2 + (v - N/2 + v_0)^2]^{1/2} \tag{5.4.7}$$

通常，根据4.2.3节所述的滤波过程，假定频率矩形的中心已经移到点$(M/2, N/2)$。因此，(u_0, v_0)的值对应移动中心。

阶数为 n 的巴特沃思陷波带阻滤波器的传递函数为：

$$H(u,v) = \frac{1}{1 + \left[\dfrac{D_0^2}{D_1(u,v)\,D_2(u,v)}\right]^n} \tag{5.4.8}$$

这里，$D_1(u,v)$和$D_2(u,v)$在式(5.4.6)和式(5.4.7)中已分别给出。高斯陷波带阻滤波器的表达式为：

$$H(u,v) = 1 - e^{-\frac{1}{2}\left[\frac{D_1(u,v)\,D_2(u,v)}{D_0^2}\right]} \tag{5.4.9}$$

有趣的是，当 $u_0 = v_0 = 0$ 时，这3个滤波器都变为高通滤波器。

正如前面的带通滤波器部分所示的，我们可以得到陷波滤波器，它能通过（而不是阻止）包含在陷波区的频率。由于这些滤波器能准确执行式(5.4.5)、式(5.4.8)和式(5.4.9)给出的陷波带阻滤波器函数的取反操作，它们的传递函数可写为：

$$H_{np}(u,v) = 1 - H_{nr}(u,v) \tag{5.4.10}$$

这里，$H_{np}(u,v)$是陷波带通滤波器的传递函数，该陷波带通滤波器与传递函数为 $H_{nr}(u,v)$的陷波带阻滤波器相对应。这作为一个练习留给读者（习题5.13），去推导刚刚讨论的与带阻滤波器相对应的陷波带通滤波器，并说明当 $u_0 = v_0 = 0$ 时，陷波带通滤波器变为低通滤波器。

例 5.8 使用陷波滤波器消除周期性噪声

图5.19(a)显示了与图4.21(a)相同的图像。在4.3.4节讨论低通滤波器时，曾指出过有更好的方法减少扫描线的影响。以下的陷波滤波方式会减少本图的噪声，而不会带来模糊。除非因为4.3节讨论过的原因，模糊是我们所希望的，否则如果找到了一个合适的滤波器，陷波滤波会更好一些。

观察图5.19(a)中噪声模式的近似水平线，希望它在频率域内的分布沿垂直轴集中。然而，没有如图5.19(b)所显示的那样明显的噪声，能够在垂直轴上有清晰的模式。但是，沿着傅里叶变换的垂直轴建立一个简单的理想陷波带通滤波器可以得到近似噪声分布，如图5.19(c)所示。噪声模式的空间表示（陷波带通滤波器结果的反变换）如图5.19(d)所示。这种噪声模式与图5.19(a)中的模式相对应。创建一个合适的陷波带通滤波器以合理的程度屏蔽噪声后，可从式(5.4.10)获得相应的陷波带阻滤波器。使用陷波带阻滤波器处理图像的效果如图5.19(e)所示。该图像相比图5.19(a)几乎见不到噪声扫描线。

5.4.4 最佳陷波滤波器

通常不会清楚地定义干扰模式。用光电扫描仪获得的图像（如太空和航空成像），有时会被扫描仪电路内的低水平信号的耦合和放大所污染。处理的图像通常包含显著的二维周期性结构，以迄今为止讨论过的最复杂的方式叠加在场景数据上。

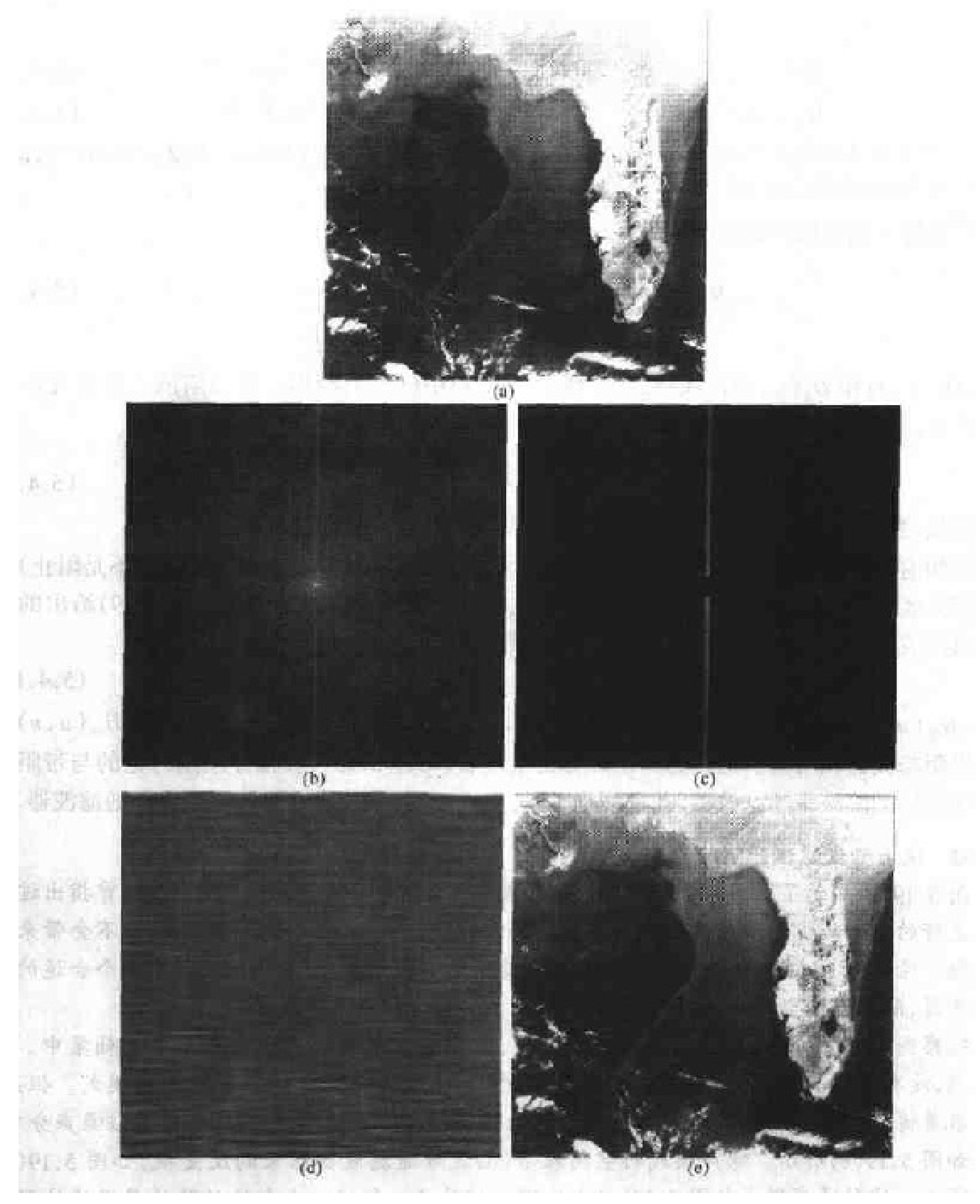


图 5.19 (a)佛罗里达和墨西哥湾的人造卫星图像(注意水平传感器扫描线),(b)(a)图的频谱,(c)叠加在(b)图的陷波带通滤波器,(d)滤波后图像的反傅里叶变换,在空间域显示噪声模式,(e)陷波带阻滤波器效果(原图像由美国国家海洋和大气局提供)

作为一个周期性图像退化的例子,图 5.20(a)显示了“水手 6 号”飞船拍摄的关于火星地形的数字图像。其干扰模式与图 5.16(a)中显示的非常相似,不过,干扰方式相当精细,因而在频率平面上更难检测。图 5.20(b)显示了所讨论图像的傅里叶频谱。其星形分量是由于扰引起的,它指出该模式包含不止一个正弦分量。

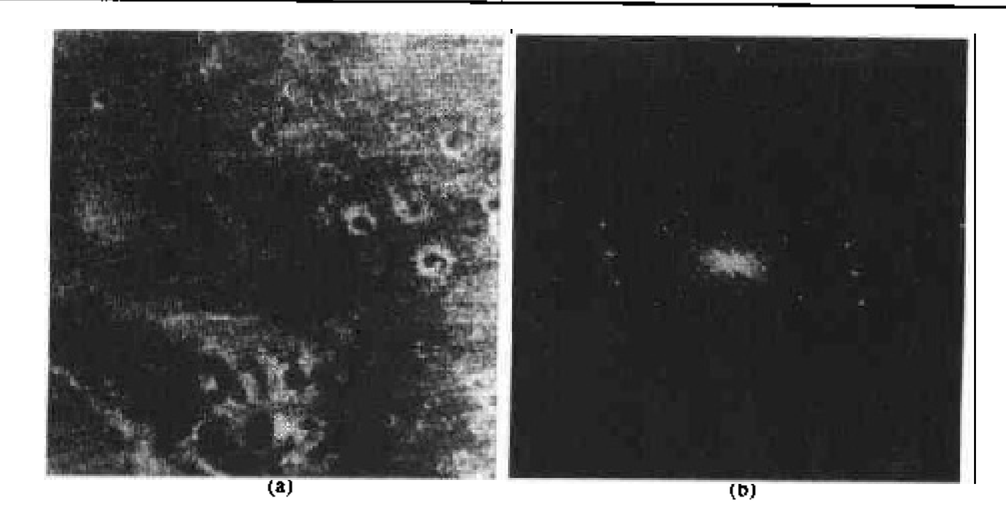


图 5.20 (a)"水手6号"飞船拍摄的火星地形图像,(b)周期干扰的傅里叶频谱(由美国国家航空和宇宙航行局提供)

当存在几种干扰时,前面介绍的方法有时就不可以采用了,因为在滤波过程中可能消除太多的图像信息(当图像为特殊的或者很难获得时,一种很不希望有的特征)。另外,干扰成分通常不是单频脉冲。相反,它们通常有宽广的携带干扰模式信息的边界。这些边界有时从正常的变换背景中不容易检测。在许多应用中,可选择的减少瑕疵影响的滤波方法非常有用。这里讨论的方法是最佳的,因为在一定意义上,它最小化了复原估计值 $\hat{f}(x,y)$ 的局部方差。

过程由两步组成,第一步屏蔽干扰的主要因素,然后,从被干扰的图像中减去一个可变的模式加权部分。虽然我们是在一个特殊应用中开发了这一过程,但是,这个基本方法仍然十分通用,而且能应用到其他复原工作中,在这些复原工作中,多周期性干扰是主要问题。

第一步是提取干扰模式的主频率成分。这可通过在每个尖峰处设一陷波带通滤波器 $H(u,v)$ 完成。如果 $H(u,v)$ 设置为只可通过与干扰模式相关的成分,那么,从 5.4.2 节和 5.4.3 节的讨论可知,干扰噪声模式的傅里叶变换由下式给出:

$$N(u,v) = H(u,v)G(u,v) \tag{5.4.11}$$

这里,如通常那样,$G(u,v)$ 仍为被污染图像的傅里叶变换。

$H(u,v)$ 的形式需要多方面判断是否有尖峰噪声干扰。为此,通常要观察显示的 $G(u,v)$ 频谱,交互地创建陷波带通滤波器。选中一个特殊滤波器之后,空间域内的相应模式就可从下式获得:

$$\eta(x,y) = \Im^{-1}\{H(u,v)G(u,v)\} \tag{5.4.12}$$

因为被污染图像假设是由未污染的图像 $f(x,y)$ 与其干扰相加形成的,若 $\eta(x,y)$ 完全已知,则如本章前面所述,从 $g(x,y)$ 减去模式得到 $f(x,y)$ 将是非常简单的事。当然,问题是,这个滤波过程通常只会得到真实模式的近似值。在 $\eta(x,y)$ 的估计中不存在的分量影响能用从 $g(x,y)$ 中减去 $\eta(x,y)$ 的加权部分得到 $f(x,y)$ 的估计值来代替最小化:

$$\hat{f}(x,y) = g(x,y) - w(x,y)\eta(x,y) \tag{5.4.13}$$

这里,如先前那样,$\hat{f}(x,y)$ 是 $f(x,y)$ 的估计值,$w(x,y)$ 已定。函数 $w(x,y)$ 称为加权函数或调制函数,此过程的目的就是选取该函数,然后以某种有意义的方法优化结果。一种方法就是选取 $w(x,y)$,使估计值 $\hat{f}(x,y)$ 在每一点 (x,y) 的指定邻域上的方差最小。

考虑点 (x,y) 的尺寸为 $(2a+1)\times(2b+1)$ 的邻域。在坐标 (x,y) 处,$\hat{f}(x,y)$ 的局部方

差可根据下面的示例估计:

$$\sigma^2(x,y) = \frac{1}{(2a+1)(2b+1)}\sum_{s=-a}^{a}\sum_{t=-b}^{b}[\hat{f}(x+s,y+t) - \bar{\hat{f}}(x,y)]^2 \quad (5.4.14)$$

这里,$\bar{\hat{f}}(x,y)$是$\hat{f}$在其邻域内的平均值,即:

$$\bar{\hat{f}}(x,y) = \frac{1}{(2a+1)(2b+1)}\sum_{s=-a}^{a}\sum_{t=-b}^{b}\hat{f}(x+s,y+t) \quad (5.4.15)$$

把式(5.4.13)代入式(5.4.14)得:

$$\begin{aligned}\sigma^2(x,y) = \frac{1}{(2a+1)(2b+1)}\sum_{s=-a}^{a}\sum_{t=-b}^{b}\{[g(x+s,y+t) \\ - w(x+s,y+t)\eta(x+s,y+t)] \\ - [\bar{g}(x,y) - \overline{w(x,y)\eta(x,y)}]\}^2\end{aligned} \quad (5.4.16)$$

假设 $w(x,y)$ 在整个邻域内保持不变,则当 $-a \leqslant s \leqslant a$ 和 $-b \leqslant t \leqslant b$ 时,给出近似式:

$$W(x+s,y+t) = w(x,y) \quad (5.4.17)$$

这一假设也可在邻域内得出如下表达式:

$$\overline{w(x,y)\eta(x,y)} = w(x,y)\bar{\eta}(x,y) \quad (5.4.18)$$

由近似值,式(5.4.16)变为:

$$\begin{aligned}\sigma^2(x,y) = \frac{1}{(2a+1)(2b+1)}\sum_{s=-a}^{a}\sum_{t=-b}^{b}\{[g(x+s,y+t) \\ - w(x,y)\eta(x+s,y+t)] \\ - [\bar{g}(x,y) - w(x,y)\bar{\eta}(x,y)]\}^2\end{aligned} \quad (5.4.19)$$

将 $\sigma^2(x,y)$最小化,解:

$$\frac{\partial\sigma^2(x,y)}{\partial w(x,y)} = 0 \quad (5.4.20)$$

对于 $w(x,y)$,解为:

$$w(x,y) = \frac{\overline{g(x,y)\eta(x,y)} - \bar{g}(x,y)\bar{\eta}(x,y)}{\overline{\eta^2}(x,y) - \bar{\eta}^2(x,y)} \quad (5.4.21)$$

要获得复原图像 $\hat{f}(x,y)$,可根据式(5.4.21)算出 $w(x,y)$,然后使用式(5.4.13)。如果 $w(x,y)$在某一邻域内被假定为常量,则不必再在图像中计算对应每一个 x,y 值的函数值。代之而来的是,在每一个非重叠邻域的一点(一般为中心点)计算 $w(x,y)$值,然后处理该邻域内包含的所有图像点。

例 5.9 最佳陷波滤波示例

从图 5.21 到图 5.23 显示了对图 5.20(a)应用前述技术处理后的效果。该图像的大小为 512×512 像素,并且 $a = b = 15$ 的邻域被选中。图 5.21 显示了被干扰图像的傅里叶谱。在这个特殊情况中,原始图像没有移到频率平面的中心,因此 $u = v = 0$ 在图 5.21 显示的变换图像的左上角。图 5.22(a)显示了 $N(u,v)$谱,这里存在尖峰噪声。图 5.22(b)显示了取 $N(u,v)$的反傅里叶变换得到的干扰模式 $\eta(x,y)$ 。注意,这个干扰模式与图 5.20(a)中存在的噪声结构很相似。最后,图 5.23 显示了应用式(5.4.13)处理后得到的图像。由于实际需要,周期干扰已被消除。

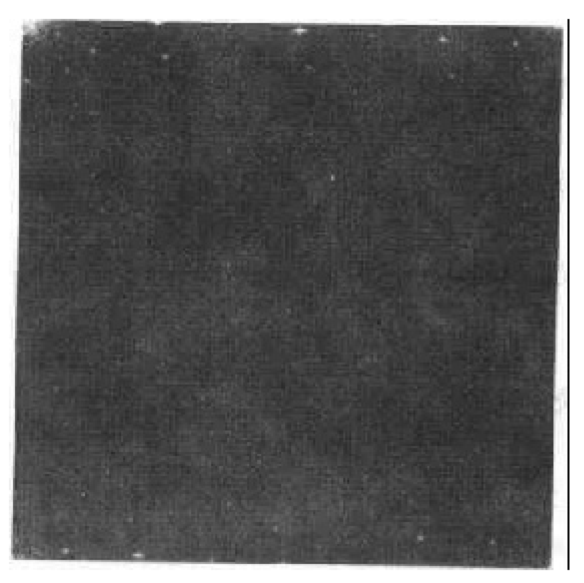


图5.21　在图5.20(a)中显示的图像的傅里叶谱(没有移动)(由美国国家航空和宇宙航行局提供)

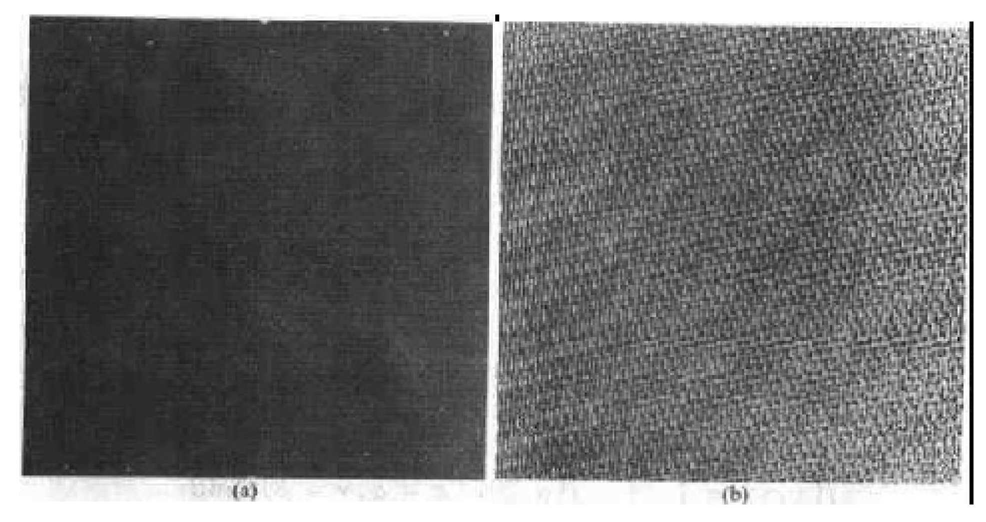


图5.22　(a)$N(u,v)$的傅里叶谱,(b)相应的噪声干扰模式$\eta(x,y)$(由美国国家航空和宇宙航行局提供)

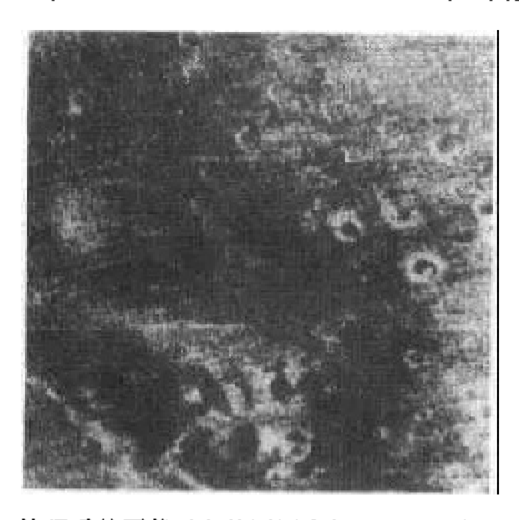


图5.23　处理后的图像(由美国国家航空和宇宙航行局提供)

5.5 线性、位置不变的退化

复原前,图5.1中的输入输出关系可以表示为:

$$g(x,y) = H[f(x,y)] + \eta(x,y) \tag{5.5.1}$$

现在,假设 $\eta(x,y) = 0$ 则 $g(x,y) = H[f(x,y)]$。基于2.6节的所述内容,如果

$$H[af_1(x,y) + bf_2(x,y)] = aH[f_1(x,y)] + bH[f_2(x,y)] \tag{5.5.2}$$

则系统 H 是一个线性系统。这里,a 和 b 是比例常数,$f_1(x,y)$ 和 $f_2(x,y)$ 是任意两幅输入图像。

若 $a = b = 1$,式(5.5.2)变为:

$$H[f_1(x,y) + f_2(x,y)] = H[f_1(x,y)] + H[f_2(x,y)] \tag{5.5.3}$$

这就是所谓的加性。这一特性简单地表明,如果 H 为线性算子,那么,两个输入之和的响应等于两个响应之和。

如果 $f_2(x,y) = 0$,式(5.5.2)变为:

$$H[af_1(x,y)] = aH[f_1(x,y)] \tag{5.5.4}$$

这称为均匀性。它表明任何与常数相乘的输入的响应等于该输入响应乘以相同的常数。即一个线性算子具有加性和均匀性。

对于任意 $f(x,y)$,α 和 β,如果

$$H[f(x-\alpha,y-\beta)] = g(x-\alpha,y-\beta) \tag{5.5.5}$$

则一个具有输入输出关系 $g(x,y) = H[f(x,y)]$ 的系统称为位置不变系统(或空间不变系统)。这个定义说明图像中任一点的响应只取决于在该点的输入值,而与该点的位置无关。

在式(4.2.33)中,离散脉冲函数定义的符号有轻微(但等效)变化,$f(x,y)$ 可描述为连续冲激函数:

$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\alpha,\beta)\delta(x-\alpha,y-\beta)\mathrm{d}\alpha\mathrm{d}\beta \tag{5.5.6}$$

事实上,这个定义中应用了位于 (x,y) 坐标处的一个单位冲激的连续变量。

现在,再次假定 $\eta(x,y)=0$,那么,把式(5.5.6)代入式(5.5.1),结果为:

$$g(x,y) = H[f(x,y)] = H\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\alpha,\beta)\delta(x-\alpha,y-\beta)\mathrm{d}\alpha\mathrm{d}\beta\right] \tag{5.5.7}$$

如果 H 是线性算子,可相加特性扩展为积分,则

$$g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H[f(\alpha,\beta)\delta(x-\alpha,y-\beta)]\mathrm{d}\alpha\mathrm{d}\beta \tag{5.5.8}$$

由于 $f(\alpha,\beta)$ 独立于 x,y,使用均匀性,可得出:

$$g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\alpha,\beta)H[\delta(x-\alpha,y-\beta)]\mathrm{d}\alpha\mathrm{d}\beta \tag{5.5.9}$$

式

$$h(x,\alpha,y,\beta) = H[\delta(x-\alpha,y-\beta)] \tag{5.5.10}$$

称为系统 H 的冲激响应。也就是说,在式(5.5.1)中,如果 $\eta(x,y)=0$,则 $h(x,\alpha,y,\beta)$ 是系统 H 对坐标 (x,y) 处强度为1的冲激的响应。在光学中,冲激为一光点,所以 $h(x,\alpha,y,\beta)$ 一般称为

点扩散函数(PSF)。这一名称来自于这样的事实,即所有物理光学系统在一定程度上模糊(扩散)光点,模糊程度由光学部件的质量决定。

将式(5.5.10)代入式(5.5.9)得:

$$g(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(\alpha,\beta)h(x,\alpha,y,\beta)\mathrm{d}\alpha\mathrm{d}\beta \tag{5.5.11}$$

这就是非常重要的第一类重叠(弗雷德霍姆)积分。这一表达式表述的结果是线性系统理论的核心。它说明,如果系统 H 对冲激函数的响应已知,则对任意输入 $f(\alpha,\beta)$ 的响应可用式(5.5.11)求得。换句话说,线性系统 H 完全可由其冲激响应来表示。

H 在位置不变的情况下,从式(5.5.5)得:

$$H[\delta(x-\alpha,y-\beta)]=h(x-\alpha,y-\beta) \tag{5.5.12}$$

式(5.5.11)简化为:

$$g(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(\alpha,\beta)h(x-\alpha,y-\beta)\mathrm{d}\alpha\mathrm{d}\beta \tag{5.5.13}$$

该式称为卷积积分。式(5.5.13)是式(4.2.30)离散卷积表达式的连续变量的等价式。这个积分告诉我们,对于任意输入 f,已知线性系统的冲激响应,就可以计算出它的响应 g。结果是冲激响应和输入函数的简单卷积。

在有加性噪声的情况下,线性退化模型[式(5.5.11)]可表示为:

$$g(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(\alpha,\beta)h(x,\alpha,y,\beta)\mathrm{d}\alpha\mathrm{d}\beta+\eta(x,y) \tag{5.5.14}$$

如果 H 是位置不变的,式(5.5.14)可表示为:

$$g(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(\alpha,\beta)h(x-\alpha,y-\beta)\mathrm{d}\alpha\mathrm{d}\beta+\eta(x,y) \tag{5.5.15}$$

噪声 $\eta(x,y)$ 的值是随机的,并假设噪声与位置无关。使用常用的卷积符号,式(5.5.15)可写为:

$$g(x,y)=h(x,y)*f(x,y)+\eta(x,y) \tag{5.5.16}$$

或者基于卷积定理,在频率域内表示如下:

$$G(u,v)=H(u,v)F(u,v)+N(u,v) \tag{5.5.17}$$

这两个表达式与式(5.1.1)和式(5.1.2)是一致的。请记住,对于离散值来说,所有的乘积是逐项相乘。例如,$H(u,v)F(u,v)$ 的 ij 项是 $H(u,v)$ 的 ij 项和 $F(u,v)$ 的 ij 项的乘积。

总之,前面所述表明,具有加性噪声的线性空间不变退化系统,可在空间域被模型化为退化函数(点扩散函数)与图像的卷积,并加上噪声。以卷积定理为基础(4.2.4 节和 4.6.4 节),同样的过程可在频率域内表示为图像和退化函数变换的乘积,然后加上噪声的变换。在频率域内工作时,可使用 4.6 节讨论的 FFT 算法。还应记住,如 4.6.3 节所述,在执行离散傅里叶变换时,需要函数延拓。

许多退化类型可近似表示为线性的位置不变过程。这一方法的优点在于广泛的线性系统理论工具对于解决图像复原问题很实用。非线性的与位置有关的技术,虽然更加普遍(通常会更加精确),但是它们会带来很多困难,常常没有解,或者解决计算问题时非常困难。这一章将集中在线性、空间不变的复原技术。由于退化模型为卷积的结果,并且图像复原需要有滤波器(该滤波器可应用于逆过程),所以术语“图像去卷积”通常用来表示线性图像复原。同样,用于复原处理的滤波器称为“去卷积滤波器”。

5.6 估计退化函数

在图像复原中,有3种主要的估计退化函数的方法:(1)观察法,(2)试验法,(3)数学建模法。在以下章节中将讨论这些方法。使用以某种方式估计的退化函数复原一幅图像的过程有时称为盲目去卷积,因为真正的退化函数很少能完全知晓。

5.6.1 图像观察估计法

假设提供了一幅退化图像,而没有退化函数 H 的知识,那么估计该函数的一个方法就是收集图像自身的信息。例如,如果图像是模糊的,可以观察包含简单结构的一小部分图像,像某一物体和背景的一部分。为了减少观察时的噪声影响,可以寻找强信号内容区。使用目标和背景的样品灰度级,可以构建一个不模糊的图像,该图像和看到的子图像有相同大小和特性。用 $g_s(x,y)$ 定义观察的子图像,用 $\hat{f}_s(x,y)$ 表示构建的子图像(现实中,它是原始图像在该区域的估计图像)。假定噪声效果可忽略,由于选择了一强信号区,根据式(5.5.17)得:

$$H_s(u,v) = \frac{G_s(u,v)}{\hat{F}_s(u,v)} \tag{5.6.1}$$

从这一函数特性,并假设位置不变,可以推出完全函数 $H(u,v)$。例如,假设 $H_s(u,v)$ 的径向曲线显现出巴特沃思低通滤波器的形状。我们可以利用这一信息在更大比例上构建一个函数 $H(u,v)$,但它有相同的形状。

5.6.2 试验估计法

如果可以使用与获取退化图像的设备相似的装置,理论上得到一个准确的退化估计是可能的。与退化图像类似的图像可以通过各种系统设置得到,退化这些图像使其尽可能接近希望复原的图像。利用相同的系统设置,由成像一个脉冲(小亮点)得到退化的冲激响应。如5.5节表明的那样,线性的空间不变系统完全由它的冲激响应来描述。一个冲激可由明亮的亮点来模拟,并使它尽可能亮以减少噪声的干扰。回顾一下,由于冲激的傅里叶变换是一个常量,由式(5.5.17)得:

$$H(u,v) = \frac{G(u,v)}{A} \tag{5.6.2}$$

这里,函数 $G(u,v)$ 与前面一样,是观察图像的傅里叶变换。A 是一个常量,表示冲激强度。图5.24显示了一个例子。

5.6.3 模型估计法

由于退化模型可解决图像复原问题,因此多年来一直在应用。在某些情况下,模型要把引起退化的环境因素考虑在内。例如,Hufnagel 和 Stanley[1964]提出的退化模型是基于大气湍流的物理特性。这个模型有一个通用公式:

$$H(u,v) = e^{-k(u^2+v^2)^{5/6}} \tag{5.6.3}$$

这里,k 是常数,它与湍流的性质有关。除了指数为5/6次方之外,这个公式与4.3.3节所述的

高斯低通滤波有相同的形式。事实上,高斯 LPF 可用来淡化模型,均匀模糊。图 5.25显示了用式(5.6.3),并且取 $k=0.0025$(在这种情况下为剧烈湍流),$k=0.001$(中等湍流)和 $k=0.00025$(轻微湍流)模拟模糊一幅图像得到的例子。所有的图像大小都为 480×480 像素。

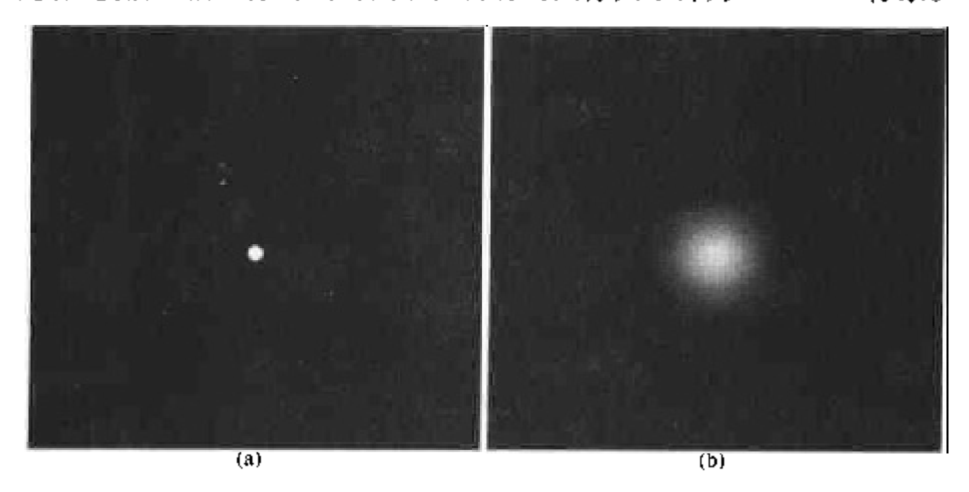


图 5.24　冲激特性的退化估计。(a) 一个亮脉冲(放大显示),(b)图像化的(退化的)冲激

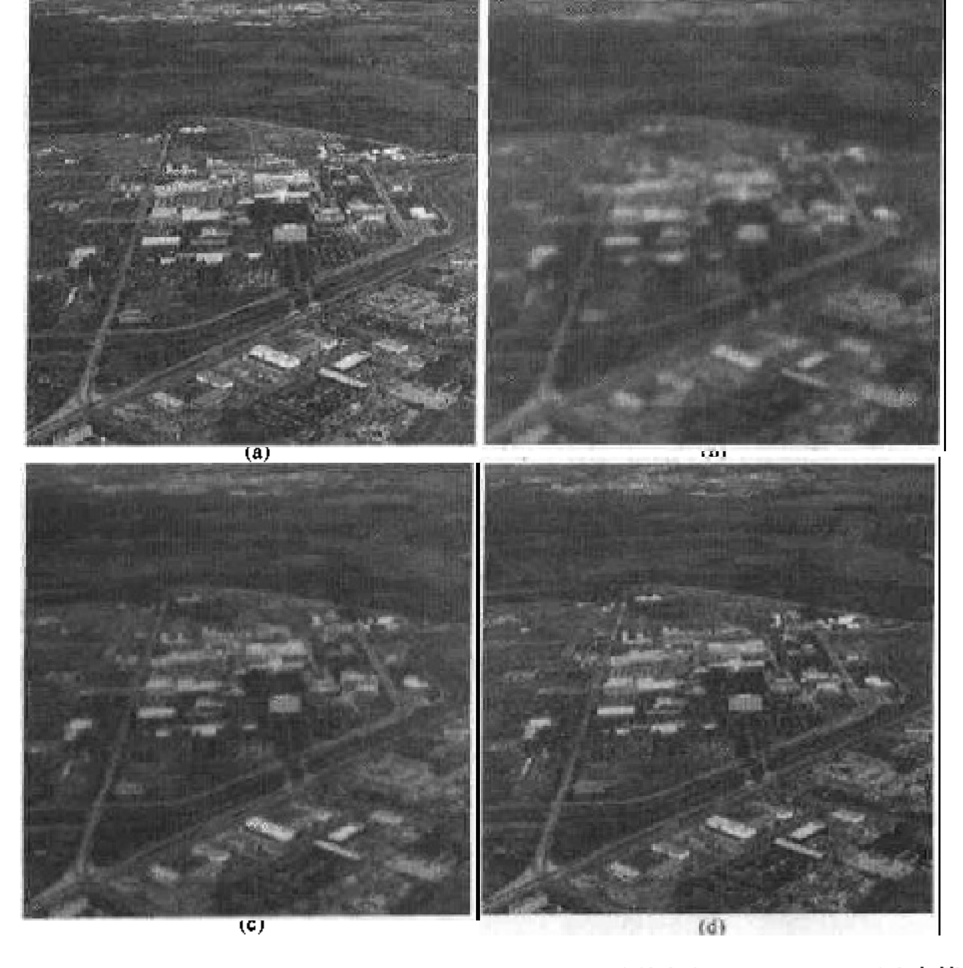


图 5.25　大气湍流模型的解释。(a)可忽略的湍流,(b)剧烈湍流,$k=0.0025$,(c)中等湍流,$k=0.001$,(d)轻微湍流,$k=0.00025$(由美国国家航空航天局提供)

模型化的另一个主要方法是从基本原理开始推导一个数学模型。通过在某些细节上处理一实例来解释这一过程,在该例中,图像获取时被图像与传感器之间的均匀线性运动模糊了。假设图像 $f(x,y)$ 进行平面运动,$x_0(t)$ 和 $y_0(t)$ 分别是在 x 和 y 方向上相应的随时间变化的运动参数。那么,在记录介质(如胶片或数字存储器)任意点的曝光总数是通过对时间间隔内瞬时曝光数的积分得到的,在该时间段,图像系统的快门是开着的。

假设快门的开启和关闭所用时间非常短,那么光学成像过程不会受到图像运动的干扰,非常完美。如果,设 T 为曝光时间,结果为:

$$g(x,y) = \int_0^T f[x - x_0(t), y - y_0(t)]\mathrm{d}t \tag{5.6.4}$$

$g(x,y)$ 为模糊的图像。

由式(4.2.3),得到式(5.6.4)的傅里叶变换为:

$$\begin{aligned} G(u,v) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\mathrm{e}^{-\mathrm{j}2\pi(ux+vy)}\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\int_0^T f[x - x_0(t), y - y_0(t)]dt\right]\mathrm{e}^{-\mathrm{j}2\pi(ux+vy)}\mathrm{d}x\mathrm{d}y \end{aligned} \tag{5.6.5}$$

改变积分顺序,式(5.6.5)可表示为:

$$G(u,v) = \int_0^T\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f[x - x_0(t), y - y_0(t)]\ \mathrm{e}^{-\mathrm{j}2\pi(ux+vy)}\mathrm{d}x\ \mathrm{d}y\right]\mathrm{d}t \tag{5.6.6}$$

外层括号内的积分项是置换函数 $f[x - x_0(t), y - y_0(t)]$ 的傅里叶变换。利用式(4.6.2)得到表达式:

$$\begin{aligned} G(u,v) &= \int_0^T F(u,v)\ \mathrm{e}^{-\mathrm{j}2\pi[ux_0(t)+vy_0(t)]}\ \mathrm{d}t \\ &= F(u,v)\int_0^T \mathrm{e}^{-\mathrm{j}2\pi[ux_0(t)+vy_0(t)]}\ \mathrm{d}t \end{aligned} \tag{5.6.7}$$

这里,最后一步根据 $F(u,v)$ 与 t 无关得出。

令

$$H(u,v) = \int_0^T \mathrm{e}^{-\mathrm{j}2\pi[ux_0(t)+vy_0(t)]}\mathrm{d}t \tag{5.6.8}$$

式(5.6.7)可表示为熟悉的形式:

$$G(u,v) = H(u,v)F(u,v) \tag{5.6.9}$$

若运动变量 $x_0(t)$,$y_0(t)$ 已知,传递函数 $H(u,v)$ 可直接由式(5.6.8)得到。假设当前图像只在 x 方向以给定的速度 $x_0(t) = at/T$ 做均匀直线运动。当 $t = T$ 时,图像由总距离 a 取代。令 $y_0(t) = 0$,式(5.6.8)导出:

$$\begin{aligned} H(u,v) &= \int_0^T \mathrm{e}^{-\mathrm{j}2\pi ux_0(t)}\ \mathrm{d}t \\ &= \int_0^T \mathrm{e}^{-\mathrm{j}2\pi uat/T}\mathrm{d}t \\ &= \frac{T}{\pi ua}\sin(\pi ua)\ \mathrm{e}^{-\mathrm{j}\pi ua} \end{aligned} \tag{5.6.10}$$

它表明,当 u 设定为 $u = n/a$ 时,H 就会变为 0,此时 n 为整数。若允许 y 分量也变化,按 $y_0 = bt/T$ 运动,则退化函数变为:

$$H(u,v)=\frac{T}{\pi(ua+vb)}\sin[\pi(ua+vb)]\,e^{-j\pi(ua+vb)} \tag{5.6.11}$$

例5.10 由于运动而引起的图像模糊

下面用式(5.6.11)说明模糊特性。图5.26(b)是由计算图5.26(a)中图像的傅里叶变换模糊的一幅图像,用式(5.6.11)中的 $H(u,v)$ 乘以变换,再做反变换。图像尺寸为688×688像素,在式(5.6.11)中使用的参数为 $a=b=0.1$ 和 $T=1$。正如在5.8节和5.9节讨论的那样,从模糊图像中复原出相应的原始图像会有某些挑战,特别是退化图像中存在噪声时。

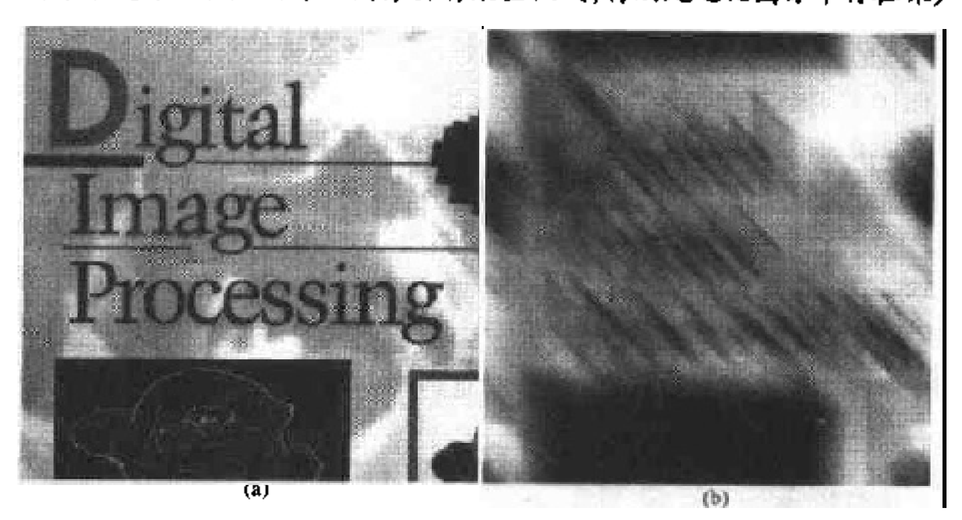


图5.26 (a)原始图像,(b)用式(5.6.11)中的函数且令 $a=b=0.1$ 和 $T=1$ 模糊的结果

5.7 逆滤波

本节的材料是研究复原由退化函数 H(由前面章节讨论的方法给出)退化的图像的第一步。最简单的方法是直接逆滤波。在该方法中,用退化函数除退化图像的傅里叶变换($G(u,v)$)来计算原始图像的傅里叶变换估计 $\hat{F}(u,v)$:

$$\hat{F}(u,v)=\frac{G(u,v)}{H(u,v)} \tag{5.7.1}$$

该式正如在式(5.5.17)中解释过的那样,是在函数的独立元素间相除。对式(5.7.1)中的 $G(u,v)$ 用式(5.5.17)的右侧替换,得到:

$$\hat{F}(u,v)=F(u,v)+\frac{N(u,v)}{H(u,v)} \tag{5.7.2}$$

这是一个有趣的表达式。它告诉我们即使知道退化函数,也不能准确地复原未退化的图像[$F(u,v)$的傅里叶反变换],因为 $N(u,v)$ 是一个随机函数,而它的傅里叶变换未知。还有更糟的情况。如果退化是零或非常小的值,$N(u,v)/H(u,v)$ 之比很容易决定 $\hat{F}(u,v)$ 的估计值。事实上,这是经常出现的情况,一会儿就将证明这一点。

一种解决退化是零或者很小值问题的途径是限制滤波的频率使其接近原点值。从式(4.2.22)知道 $H(0,0)$ 等于 $h(x,y)$ 的平均值,而且常常是 $H(u,v)$ 在频域的最高值。所以,通过将频率限制为接近原点进行分析,就减少了遇到零值的几率。这种方法在下面的例子中再进行解释。

例 5.11 逆滤波

图 5.25(b)中是通过对退化图像的退化函数精确取反，并使用式(5.7.1)反向滤波的图像。其所用的退化函数是：

$$H(u,v)=e^{-k[(u-M/2)^2+(v-N/2)^2]^{5/6}}$$

其中 $k=0.0025$。$M/2$, $N/2$ 常数是偏移量值；它们可以中心化函数，与傅里叶变换中心化相对应，这个问题在前一章曾多次讨论。在这个例子中，$M=N=480$。因为高斯形函数没有零值，所以在这里没有什么关系。然而，尽管如此，退化值会变得非常小，以至于全逆滤波的结果[图 5.27(a)]变得毫无用处。产生这个失败结果的原因正如式(5.7.2)所讨论的。

图 5.27(b)到(d)显示了截止 $G(u,v)/H(u,v)$ 值的结果，半径分别为 40，70，85。这是通过对这个比值应用一个阶数为 10 的巴特沃思低通滤波器来实现的。它提供了对于希望的半径的较尖锐(但平滑)过渡。半径在 70 左右的时候视觉效果最好[图 5.27(c)]，70 以下开始变得模糊，正如图 5.27(b)显示的那样，这里，是用半径 40 得到的结果。高于 70，图像开始退化，如图 5.27(d)显示的那样，它是用半径 85 得到的结果。这个图看起来几乎像是在一个噪声“窗帘”后观看的，噪声明显影响了结果。进一步增加半径值，使图像越来越像图 5.27(a)。

这个例子的结果说明一般直接逆滤波的性能是较差的，本节的主题就是讨论怎样改进直接逆滤波的性能。

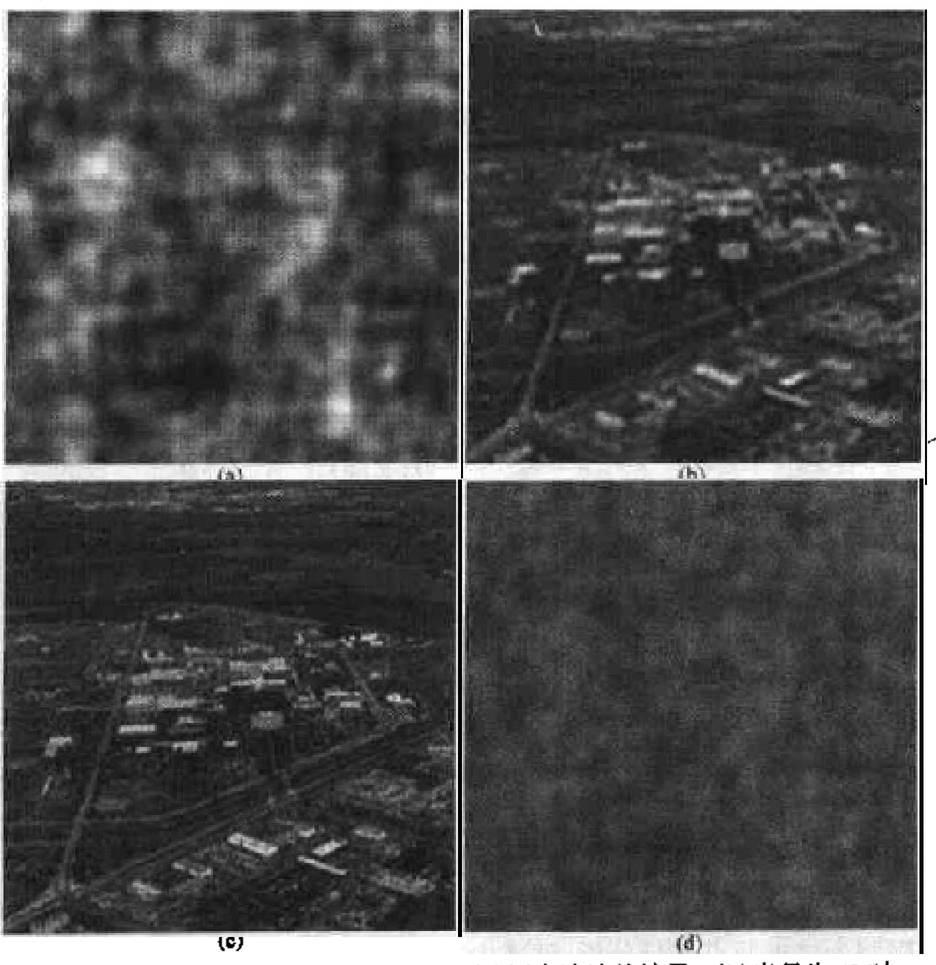


图 5.27 用式(5.7.1)复原图 5.25(b)。(a)用全滤波的结果，(b)半径为 40 时截止 H 的结果，(c)半径为70时的结果，(d)半径为85时的结果

5.8 最小均方误差滤波(维纳滤波)

前面讨论的逆滤波并没有清楚地说明怎样处理噪声,本节讨论综合了退化函数和噪声统计特征两个方面进行复原处理的方法。方法建立在认为图像和噪声是随机过程的基础上,而目标是找一个未污染图像 f 的估计值 $\hat{f}$,使它们之间的均方误差最小。误差度量由下式给出:

$$e^2 = E\{(f - \hat{f})^2\} \tag{5.8.1}$$

$E\{\cdot\}$ 是宗量的期望值。这里假定:噪声和图像不相关;其中一个有零均值;估计的灰度级是退化图像灰度级的线性函数。在这些条件下,式(5.8.1)中误差函数的最小值在频域用下列表达式计算:

$$\begin{aligned}\hat{F}(u,v) &= \left[\frac{H^*(u,v)S_f(u,v)}{S_f(u,v)|H(u,v)|^2 + S_\eta(u,v)}\right]G(u,v)\\ &= \left[\frac{H^*(u,v)}{|H(u,v)|^2 + S_\eta(u,v)/S_f(u,v)}\right]G(u,v)\\ &= \left[\frac{1}{H(u,v)}\frac{|H(u,v)|^2}{|H(u,v)|^2 + S_\eta(u,v)/S_f(u,v)}\right]G(u,v)\end{aligned} \tag{5.8.2}$$

这里,我们应用了这样一个事实:一个复数量与它的共轭的乘积等于复数量幅度的平方。这个结果就是众所周知的维纳滤波,是 N. Wiener[1942]首次提出的概念。由括号里边的项组成的滤波器通常还叫做最小均方误差滤波器,或最小二乘方误差滤波器。本章最后列出了一些参考资料,包括了维纳滤波器的详细推导过程。注意式(5.8.2)的第一行,维纳滤波器没有逆滤波中退化函数为零的问题,除非对于相同的 u,v 值,$H(u,v)$ 和 $S_\eta(u,v)$ 都是0。

式(5.8.2)中的各项如下所示:

$H(u,v)$ = 退化函数

$H^*(u,v) = H(u,v)$ 的复共轭

$|H(u,v)|^2 = H^*(u,v)H(u,v)$

$S_\eta(u,v) = |N(u,v)|^2$ = 噪声的功率谱[见式(4.2.20)]

$S_f(u,v) = |F(u,v)|^2$ = 未退化图像的功率谱

与前面一样,$H(u,v)$ 是退化函数的变换,而 $G(u,v)$ 是退化图像的变换,在空间域被复原的图像由频率域估计值 $\hat{F}(u,v)$ 的傅里叶反变换给出。注意,如果噪声是零,噪声功率谱消失,并且维纳滤波退化为逆滤波。

当处理白噪声时,谱 $|N(u,v)|^2$ 是一个常数,大大简化了处理过程。然而,未退化图像的功率谱很少是已知的。当这些值未知或不能估计时,经常使用的方法是用下面的表达式近似:

$$\hat{F}(u,v) = \left[\frac{1}{H(u,v)}\frac{|H(u,v)|^2}{|H(u,v)|^2 + K}\right]G(u,v) \tag{5.8.3}$$

这里,k 是一个特殊常数。下面的例子说明了表达式的应用。

例5.12 逆滤波和维纳滤波的比较

图5.28说明了与直接逆滤波相比维纳滤波的能力。图5.28(a)是图5.27(a)所示的全逆滤波的结果。类似地,图5.28(b)是图5.27(c)的半径受限的逆滤波结果。这些图像为方便比较重新在这里列出。图5.28(c)显示了用式(5.8.3)通过例5.11中使用的退化函数

得到的结果。K 值交互式的选择是为了寻找最好的视觉效果,在这个例子中与直接逆滤波方法相比较,维纳滤波显然更好。比较图 5.25(a)和图 5.28(c),会发现维纳滤波的结果非常接近原始图像。

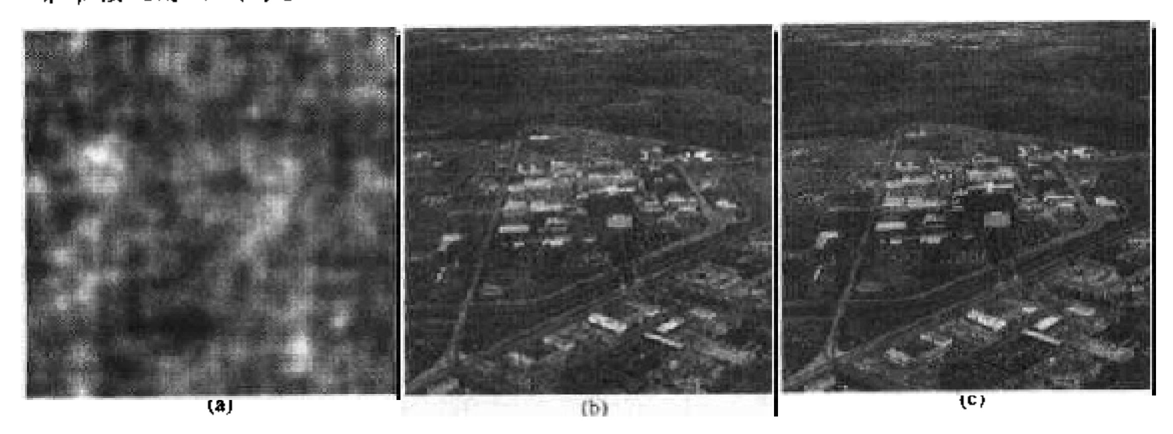


图 5.28 逆滤波和维纳滤波的比较。(a)图 5.25(b)的全逆滤波结果,(b)半径受限的逆滤波结果,(c)维纳滤波的结果

例 5.13 维纳滤波的进一步比较

图 5.29 的第一行从左到右分别是被均值为 0 和方差为 650 的加性高斯噪声严重污染的图 5.26(b)的模糊图像,及直接逆滤波和维纳滤波的结果。用式(5.8.3)的维纳滤波,通过例 5.10 的 $H(u,v)$ 和迭代地选择 K 可以得到最好的视觉效果。如期望的那样,逆滤波产生了一个不太适用的图像,注意,逆滤波图像中的噪声非常强,它的结构属于去模糊滤波。维纳滤波的结果并不完美,但它给了我们一些关于图像内容的线索。要辨认图中的文字还是有点难度。

图 5.29 的第二行显示了同样的序列,但噪声幅度的方差水平减少了一个数量级。这对逆滤波没有多少影响,但维纳滤波的效果明显改进了,图中的文字容易辨认多了。图 5.29 的第三行相对于第一行噪声方差减少了 5 个数量级。事实上,图 5.29(g)的图像已没有可见的噪声。在这种情况下,逆滤波的结果相当有趣。噪声仍很明显,但是文字能通过噪声"窗帘"看见。这是关于式(5.7.2)的一个很好的例子。换句话说,正如从图 5.29(h)可明显看到的,逆滤波消除图像模糊的能力是很强的。然而,噪声还是支配着结果,字符仍有一点模糊。图 5.29(i)的维纳滤波结果是最好的,非常接近原图。这些是使用维纳滤波的典型结果,只要可以得到退化函数的合理估计。

5.9 约束最小二乘方滤波器

已知退化函数 H 的某些知识的问题对于本章讨论的方法很常见。然而,维纳滤波却存在一些其他困难:未退化图像和噪声的功率谱必须是已知的。在前面的章节中已显示出,使用式(5.8.3)近似的方法能得到很好的结果。但是功率谱比的常数估计一般还是没有合适的解。

本节讨论的方法只要求噪声方差和均值的知识。正如在 5.2.4 节讨论的那样,那些参数经常能从一个给定的退化图像计算出来,这是一个很重要的优点。另一个不同是,维纳滤波建立在最小化统计准则的基础上,在平均意义上它是最优的。本节所提供的算法具有显著的特

点，对于处理的每一幅图像它都能产生最优结果。当然，这些理论上满足的最优准则与动态的视觉感知并没有关系。因此，算法的选择往往由结果图像的感知视觉质量决定(至少部分是这样)。

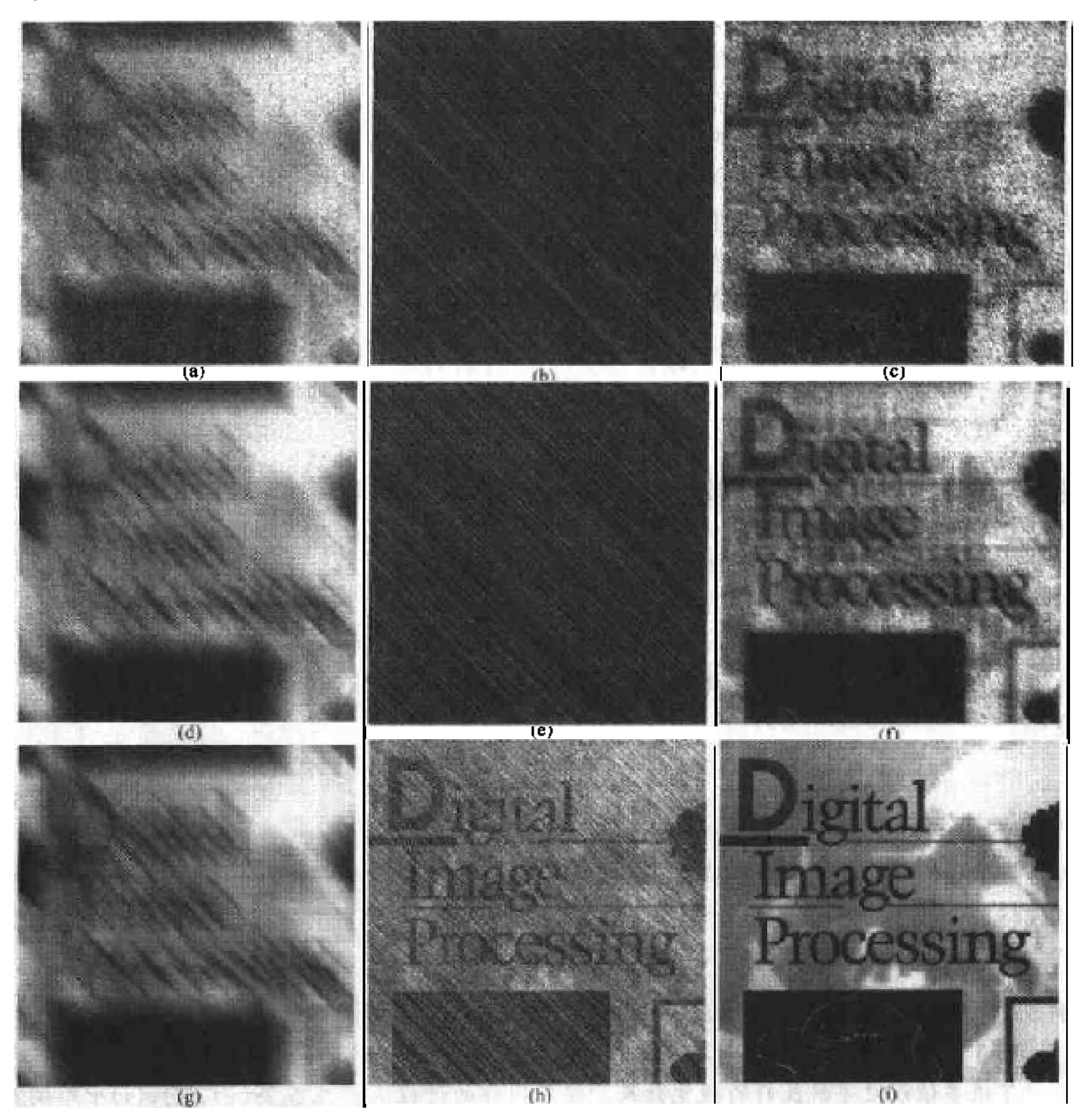


图5.29 (a)由运动模糊及加性噪声污染的图像，(b)逆滤波的结果，(c)维纳滤波的结果，(d)～(f)同样的序列，但是噪声幅度的方差小了一个数量级，(g)～(i)同样的序列，但是噪声方差比(a)小5个数量级，注意，在(h)中，去模糊图像透过噪声"窗帘"清晰可见

通过使用式(4.2.30)给出的卷积定义，把式(5.5.16)表达为下式形式：

$$\mathbf{g} = \mathbf{Hf} + \boldsymbol{\eta} \tag{5.9.1}$$

例如，设 $g(x,y)$的尺寸为 $M \times N$。可以用 $g(x,y)$第一行的图像元素构成向量 $\mathbf{g}$ 的第一组 N 个元素，下一组 N 个元素用第二行构成，以此类推。结果向量是 $MN \times 1$ 维。这也同样是 $\mathbf{f}$ 和 $\boldsymbol{\eta}$ 的维数，因为这些向量是以同样的方法构成的。$\mathbf{H}$ 矩阵有 $MN \times MN$ 维。它的元素由

式(4.2.30)的卷积元素给出。

把图像复原问题简化为矩阵处理的结论看起来很合理。遗憾的是,并不是这么简单。例如,假定处理一个中等大小的图像,大小为 $M=N=512$,那么式(5.9.1)的向量为 262 144× 1维,矩阵 **H** 为 262 144 × 262 144 维。计算这样尺寸的向量和矩阵并不是一件简单的任务。由于 **H** 对噪声高度敏感,问题被进一步复杂化了(在前面两章中,对于噪声的影响我们已了解到,对此不会感到奇怪)。然而,公式化矩阵形式的复原问题要做复原技术的简化推导。

尽管我们不能完全地推导约束最小二乘方法,但这种方法在矩阵表示法中的确有它的根源。在本章的最后会给出参考资料,指出哪里有详细的推导。该方法的核心是 **H** 对噪声的敏感性问题。减少噪声敏感性问题的一种方法以平滑措施的最佳复原为基础,如一幅图像的二阶导数("拉普拉斯变换"这位"老朋友")。意味深长的是,复原必须被手边的参量约束。因此,我们期望的是找一个最小的准则函数 C,定义如下:

$$C=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left[\nabla^2 f(x,y)\right]^2 \tag{5.9.2}$$

约束为:

$$\|\mathbf{g}-\mathbf{H}\hat{\mathbf{f}}\|^2=\|\boldsymbol{\eta}\|^2 \tag{5.9.3}$$

这里,$\|\mathbf{w}\|^2\triangleq\mathbf{w}^T\mathbf{w}$ 是欧几里得矢量范数[①],$\hat{f}$ 是未退化图像的估计值。拉普拉斯算子∇^2在式(3.7.1)中定义。

这个最佳化问题的频域解决方法由下面的表达式给出:

$$\hat{F}(u,v)=\left[\frac{H^*(u,v)}{|H(u,v)|^2+\gamma|P(u,v)|^2}\right]G(u,v) \tag{5.9.4}$$

这里,γ 是一个参数,它必须被调整以使式(5.9.3)满足条件,$P(u,v)$是函数 $p(x,y)$的傅里叶变换。

$$p(x,y)=\begin{bmatrix}0 & -1 & 0\\ -1 & 4 & -1\\ 0 & -1 & 0\end{bmatrix} \tag{5.9.5}$$

在 3.7.2 节介绍拉普拉斯算子时曾见过这个函数。$p(x,y)$和所有其他相关的空间域函数,在用式(5.9.4)计算它们的傅里叶变换之前要用零进行适当的延拓,这正如 4.6.3 节讨论的那样。注意,当 γ 为 0 时,式(5.9.4)变为逆滤波。

例 5.14 维纳和最小二乘方滤波的比较

图 5.30 显示了用约束最小二乘方滤波处理图 5.29(a),(d)和(g)的结果,其中,γ 被选择了很多值以便寻找最好的视觉效果。使用同样的过程来产生 5.29(c),(f),(i)中维纳滤波的结果。比较约束最小二乘方滤波和维纳滤波的结果,我们发现前者对于高、中噪声情况下的处理稍有优势,而在低噪条件下两者差不多。当人为地选择参数以取得更好的视觉结果时,约束最小二乘方滤波器有可能要比维纳滤波器更好。参数 γ 在式(5.9.4)中是一个标量,K 在式(5.8.3)中是两个未知频域函数的近似比,比值很少是一个常数。所以,人为选择 γ 将得出未退化图像更准确的估计。

---

① 回顾一个带有 n 个元素的向量 $\mathbf{w}$,$\mathbf{w}^T\mathbf{w}=\sum_{k=1}^{n}w_k^2$,这里,$w_k$ 是 $\mathbf{w}$ 的第 k 个元素。

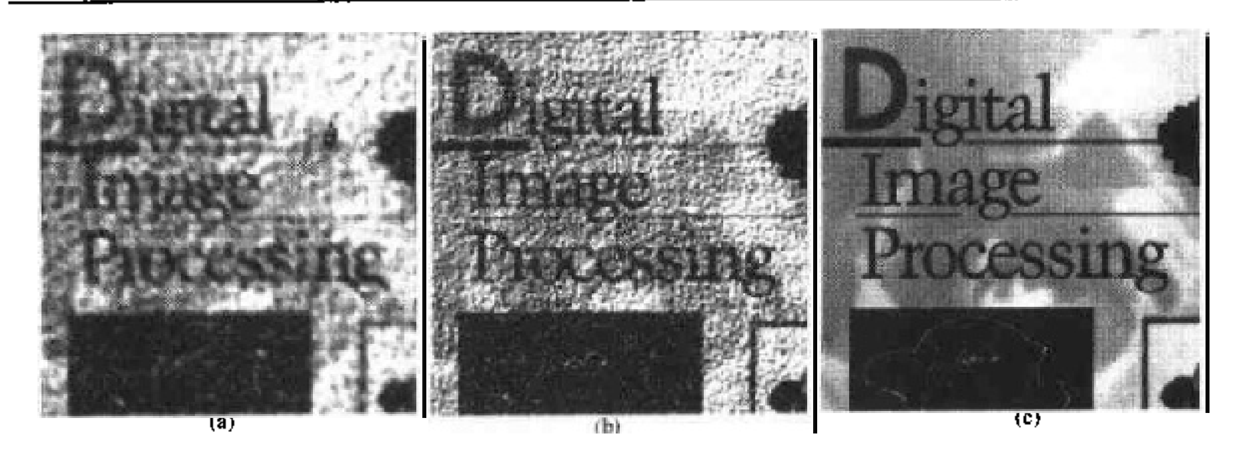


图5.30 约束最小二乘方滤波的结果，用图5.29(c)，(f)和(i)中的结果分别与维纳滤波(a)，(b)，(c)相比较

如前面的例子所示，有可能交互地调整参量 γ，直到取得可接受的结果。然而，如果对最优化感兴趣，那么参量 γ 必须调整，以便满足式(5.9.3)的约束。迭代计算 γ 遵循下面的过程。

定义一个"残数"向量 $\mathbf{r}$：

$$\mathbf{r} = \mathbf{g} - \mathbf{H}\hat{\mathbf{f}} \tag{5.9.6}$$

因为由式(5.9.4)的解可知，$\hat{F}(u,v)$(即隐含的 $\hat{\mathbf{f}}$)是 γ 的函数，所以 $\mathbf{r}$ 也是该参数的函数。如下所示(Hunt[1973])：

$$\begin{aligned}\varphi(\gamma) &= \mathbf{r}^T\mathbf{r} \\ &= \|\mathbf{r}\|^2\end{aligned} \tag{5.9.7}$$

它是 γ 的单调递增函数。这里要做的是调整 γ，以便：

$$\|\mathbf{r}\|^2 = \|\boldsymbol{\eta}\|^2 \pm a \tag{5.9.8}$$

这里，a 是一个精确的因数。在式(5.9.6)看来，如果 $\|\mathbf{r}\|^2 = \|\boldsymbol{\eta}\|^2$，那么式(5.9.3)的约束就完全满足了。

因为 $\phi(\gamma)$ 是单调的，寻找满足要求的 γ 值并不难。一种途径是：

1. 指定初始 γ 值
2. 计算 $\|\mathbf{r}\|^2$
3. 如果式(5.9.8)满足，则停止；否则如果 $\|\mathbf{r}\|^2 < \|\boldsymbol{\eta}\|^2 - a$，增加 γ，或者如果 $\|\mathbf{r}\|^2 > \|\boldsymbol{\eta}\|^2 + a$，则减少 γ，并回到第2步。使用 γ 的新值，通过式(5.9.4)重新计算最佳估计值 $\hat{F}(u,v)$。

其他的方法，譬如牛顿-拉弗森算法，也可用于改进收敛速度。

为了使用这一算法，需要量化 $\|\mathbf{r}\|^2$ 和 $\|\boldsymbol{\eta}\|^2$ 的值。要计算 $\|\mathbf{r}\|^2$，从式(5.9.6)得：

$$R(u,v) = G(u,v) - H(u,v)\hat{F}(u,v) \tag{5.9.9}$$

由此，可以通过计算 $R(u,v)$ 的傅里叶反变换得到 $r(x,y)$。然后，

$$\|\mathbf{r}\|^2 = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} r^2(x,y) \tag{5.9.10}$$

计算 $\|\boldsymbol{\eta}\|^2$，可以得出重要的结果。首先，对于整幅图像上的噪声方差，可以使用取样平均的方法估计：

$$\sigma_\eta^2 = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[\eta(x,y) - m_\eta]^2 \tag{5.9.11}$$

这里,

$$m_\eta = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\eta(x,y) \tag{5.9.12}$$

是样本均值。参考式(5.9.10),注意到式(5.9.11)的双求和等于 $\|\boldsymbol{\eta}\|^2$。这给出了如下表达式:

$$\|\boldsymbol{\eta}\|^2 = MN[\sigma_\eta^2 - m_\eta] \tag{5.9.13}$$

这是非常有用的结果,它告诉我们,可以仅仅用噪声均值和方差的知识执行最佳复原算法。这些值并不难估计(见 5.2.4 节),这时要假定噪声和图像灰度值不相关。这是本章讨论的所有方法的一个基本假设。

例 5.15 最佳约束最小二乘方的迭代估计

图 5.31(a)显示了使用刚才描述的算法,为复原图 5.25(b)估计最佳滤波器得到的结果。所用的初始值 γ 是 10^{-5},调整 γ 的校正因子是 10^{-6},a 是 0.25。用于产生图 5.25(a)的指定噪声参数相同:噪声的方差是 10^{-5},均值为 0。复原结果几乎与图 5.28(c)一样好,它是由维纳滤波得到的结果,其中,人为调整 K 以产生最好的视觉效果。图 5.31(b)显示了使用估计错误的噪声参数时会发生的情况。在这种情况下,指定的噪声方差为 10^{-2},并且均值保留为 0。其结果有相当大的模糊。

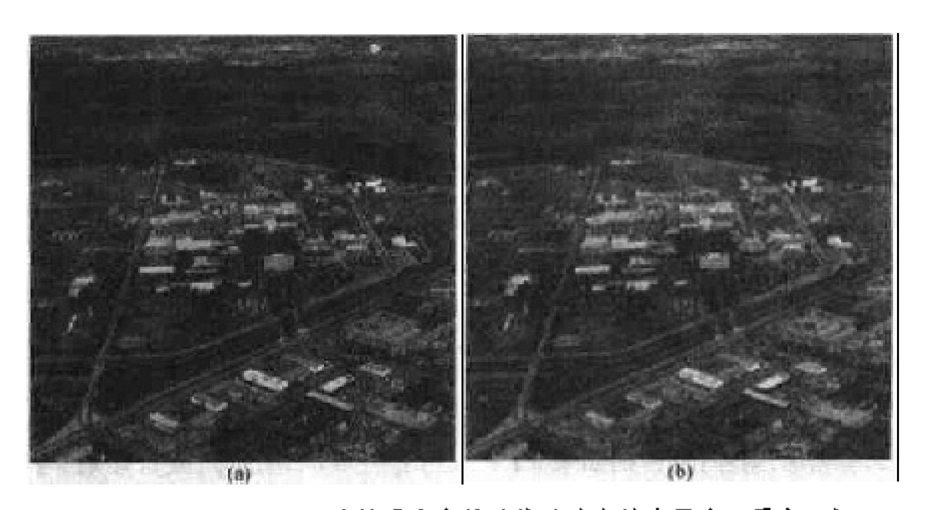


图 5.31 (a)用正确的噪声参数迭代地确定约束最小二乘方,对图5.16(b)进行复原,(b)用错误的噪声参数得到的结果

正如本章开始指出的那样,约束最小二乘方意义下的最佳复原在视觉效果上并不意味着最好。根据退化和噪声的性质及大小,算法中交互地确定最佳估计的其他参数,在最终结果中也起到很重要的作用。通常,自动确定复原滤波器相比人为调整滤波器参数的复原结果要差。特别是约束最小二乘方滤波器完全由单一的标量参数来决定时更是如此。

5.10 几何均值滤波

在5.8节中讨论了维纳滤波器,可以把它稍微加以普遍化。普遍化时采用几何均值滤波器的形式.

$$\hat{F}(u,v)=\left[\frac{H^*(u,v)}{|H(u,v)|^2}\right]^{\alpha}\left[\frac{H^*(u,v)}{|H(u,v)|^2+\beta\left[\frac{S_\eta(u,v)}{S_f(u,v)}\right]}\right]^{1-\alpha}G(u,v)\quad(5.10.1)$$

α,β是正的实常数。几何均值滤波器由两个括号内的幂次分别为α和$1-\alpha$的表达式组成。

当$\alpha=1$时,滤波退化为逆滤波;当$\alpha=0$时,变成所谓的参数维纳滤波;当$\beta=1$时,退化为标准的维纳滤波。如果$\alpha=1/2$,滤波器变成相同次幂的两个量的积,这是几何均值的定义,这样就给出了这种滤波器的命名。当$\beta=1$,α减到1/2以下时,滤波的性能越来越接近逆滤波。当α增加到1/2以上时,越来越接近维纳滤波、当$\alpha=1/2,\beta=1$时,滤波器还是通常的谱均衡滤波器。式(5.10.1)对实现复原滤波器非常有用,因为它确实表示了合并为一个表达式的滤波器族。

5.11 几何变换

在本章的最后介绍一下用于图像复原的几何变换。不像迄今为止讨论过的技术,几何变换可在一幅图像中的像素间修改空间联系。几何变换通常又叫做橡皮片变换,因为它们可以被看做在一片橡皮片上打印图像,然后根据预先确定的规则拉伸这个橡皮片的过程。

在数字图像处理中,几何变换由两个基本操作组成:(1)一个空间变换,它定义了图像平面上像素的重新安排;(2)灰度级插补,它处理空间变换后图像中像素灰度级的赋值。在下面的部分讨论这些基础概念和它们在图像复原中的应用。

5.11.1 空间变换

假设一幅图像f,像素点坐标为(x,y),经过几何失真产生了一幅图像g,像素点坐标为(x',y')。这个变换可以表示为:

$$x'=r(x,y)\quad(5.11.1)$$

$$y'=s(x,y)\quad(5.11.2)$$

这里,$r(x,y)$和$s(x,y)$是空间变换,产生了几何失真图像$g(x',y')$。例如,如果$r(x,y)=x/2$,$s(x,y)=y/2$,则失真后的图像只是简单地在两个空间方向上将$f(x,y)$的尺寸收缩为一半。

如果$r(x,y)$和$s(x,y)$在解析分析中为已知,理论上可以用相反的变换从失真图像$g(x',y')$复原$f(x,y)$。然而,在实践中,公式化一个解析函数$r(x,y)$和$s(x,y)$的集合是不可能的,这些解析函数描述了整个图像平面上的几何失真过程。最常用的克服这一困难的方法是用"连接点"表达像素的空间重定位,这些点是像素的子集,它们在输入(失真的)和输出(校正的)图像中的位置是精确已知的。

图5.32显示了在失真和相应的校正图像中的四边形区域,四边的顶点是相应的"连接点"。假设四边形区域中的几何变形过程用双线性方程对来建模,即:

$$r(x,y) = c_1x + c_2y + c_3xy + c_4 \tag{5.11.3}$$

$$s(x,y) = c_5x + c_6y + c_7xy + c_8 \tag{5.11.4}$$

从式(5.11.1)和式(5.11.2),可知:

$$x' = c_1x + c_2y + c_3xy + c_4 \tag{5.11.5}$$

$$y' = c_5x + c_6y + c_7xy + c_8 \tag{5.11.6}$$

因为总共有8个已知的"连接点",这些方程可以解出8个系数,$c_i, i = 1,2,\cdots,8$。这些系数构成了用于变换四边形区域内所有像素的几何失真模型,该四边形是由导出系数的"连接点"定义的。通常,需要足够多的连接点以产生覆盖整个图像的四边形集,每一个四边形都有它自己的系数集。

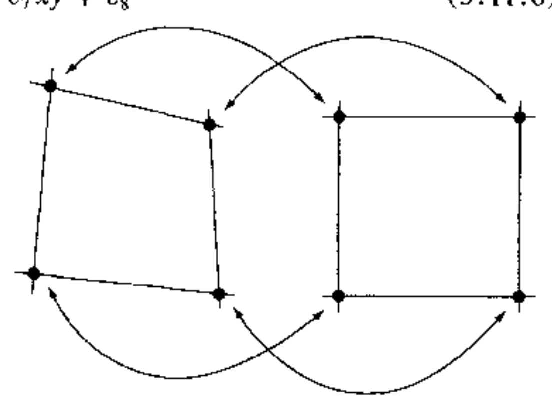


图5.32 两图像区相应的"连接点"

一旦有了系数,产生校正(即,复原)图像就不困难了。如果想找到非失真图像在任意点(x_0, y_0)的值,需要简单地知道$f(x_0, y_0)$在非失真图像中的什么地方被映射。为此,可以把(x_0, y_0)代入式(5.11.5)和式(5.11.6)得到几何失真坐标(x_0', y_0')。在无失真图像中被映射到(x_0', y_0')点的值是$g(x_0', y_0')$。这样,简单地令$\hat{f}(x_0, y_0) = g(x_0', y_0')$,就得到了复原图像的值。例如,为产生$\hat{f}(0,0)$,可以把$(x,y) = (0,0)$代入式(5.11.5)和式(5.11.6),得到坐标对(x', y')。然后,令$\hat{f}(0,0) = g(x', y')$,这里,x'和y'是刚才得到的坐标值。下一步,把$(x,y) = (0,1)$代入式(5.11.5)和式(5.11.6),从这些公式得到另一对值(x', y'),并对于坐标值令$f(0,1) = g(x', y')$。该过程逐像素和逐行地继续直到得到整个阵列,其尺寸不大于图像g的尺寸。按列扫描(而不是行)将得到相同的结果。并且,为了使用合适的系数,需要系统的记录过程,以追踪在给定像素位置的四边形。

连接点根据其应用可用多种不同技术建立。例如,一些图像生成系统有物理的人为缺陷(如金属点),镶嵌在图像传感器上。这产生了一个已知点集(叫做网状标记),在图像被获取时直接镶在图像上。如果图像由于一些其他处理(如图像显示或图像重建处理)失真了,图像可以用刚才讨论的技术校正。

5.11.2 灰度级插补

前面一节讨论的方法遍历坐标(x,y)的整数值得到复原图像$\hat{f}(x,y)$。然而,根据系数c_i的值,式(5.11.5)、式(5.11.6)能产生非整数的x', y'值。因为失真图像g是数字的,它的像素值只定义在整数坐标。因此,对x', y'用非整数值会导致一个到g位置的映射,在这些位置没有灰度定义,所以有必要基于整数坐标的灰度值去推断那些位置的灰度值。用于完成该任务的技术叫做灰度级插补。

最简单的灰度级插补是最近邻域法。这种方法也叫做零阶内插,如图5.33所示。该图显示了:(1)借助于式(5.11.5)和式(5.11.6)把整数坐标(x,y)映射到分数坐标(x', y');(2)与(x', y')相邻的最接近整数坐标的选择;(3)为这些最接近的坐标赋以位于(x,y)处的像素灰度值。虽然最近邻域内插实现上很简单,但这个方法经常有产生不希望的人为疵点的缺点,如

高分辨率图像直边的扭曲。可以用更完善的技术得到较平滑的结果，例如，立方卷积内差。这种技术为了在任何希望的点得到较平滑的灰度估计，通过大得多的邻域（比如16）拟合$\sin(z)/z$的表面。要求更平滑近似的典型应用领域是三维图形学（Watt[1993]）和医学成像（Lehman等[1999]）。更平滑的近似所付出的代价是附加的计算开销。对于通常目的的图像处理，双线性内插法很实用，该方法用4个最近邻点。这个方法很简单；因为非整数坐标对(x', y')的4个整数最近邻点的灰度级都是已知的，定义在这些坐标上的灰度级$v(x', y')$用下面的关系从它的邻点值被插补。

$$v(x', y') = ax' + by' + cx'y' + d \tag{5.11.7}$$

这里，4个系数很容易从用(x', y')4个已知邻点写出的含4个未知数的4个方程确定。当确定这些系数后，计算$v(x', y')$，并且这个值被赋给$f(x, y)$中的位置，这就是到位置(x', y')的空间映射。用图5.33可视化这一过程很容易。一个例外是，不是使用(x', y')最近邻点的灰度级，而是实际上在(x', y')点内插一个值，并用这个值在(x, y)点赋灰度级。

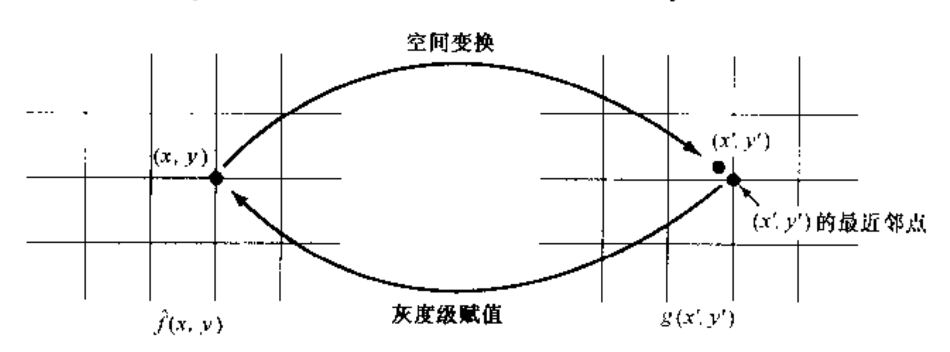


图5.33　基于最近邻点概念的灰度级插补

例5.16　几何变换的说明

图5.34(a)显示了一幅具有25个等间隔连接点（图像中重点加强的点以增强可见性）的图像。图5.34(b)显示了引起几何失真的连接点安排。结合式(5.11.5)和式(5.11.6)并参考讨论过的过程，这些方程的系数是从未失真坐标到失真坐标的映射。一旦系数已知，我们就有了一个模型，可以畸变一幅图像（为了说明）或者可以复原一幅图像，该图像是在由一组系数定义的条件下的几何失真。

假设要失真图5.34(a)的图像。可以简单地将该幅图像的任何像素(x_0, y_0)的值代入式(5.11.5)和式(5.11.6)，产生相应的坐标(x_0', y_0')，并四舍五入为最接近的整数值。失真图像在那一点的值由$g(x'_0, y'_0) = f(x_0, y_0)$给出，或者可以用$(x_0, y_0)$邻域中的$f$值的灰度内插。这是描述过的关于式(5.11.5)和式(5.11.6)的相同处理过程。这里简单地反向应用它。

用刚才讨论的方法失真图5.34(a)的结果示于图5.34(c)中。这里，应用了最近邻点灰度赋值方法。注意，这是不成功的严重失真。如果这是给定的图像，可以用讨论过的与式(5.11.5)和式(5.11.6)相联系的方法和在5.11.2节讨论过的灰度赋值技术。这一过程的结果示于图5.34(d)，并再次使用了最近邻点灰度赋值方法。注意，几何校正是可接受的，但是在灰度级赋值上有大量明显的错误，特别是沿着灰和黑区域间的边界处。图5.34(e)和(f)显示了同样的实验序列，但是使用了双线性内插法。可以看到在灰和黑

区域间的边界处的改进特别明显。

刚才讨论的图像是有规律的，并且在尖锐的边界中没有多少灰度级，几乎任何类型的几何失真都会导致明显的退化。当图像有较多的纹理时，几何校正错误变得不太明显了。例如，考虑图 5.35。图 5.35(b)是图 5.35(a)以图 5.34(e)相同的方式几何失真的结果。图 5.35(b)的这一失真近于不可见。图 5.35(a)和(b)间的差别很重要，如图 5.35(c)所示的差别图像。它们是不可见的，因为在图像中有纹理的变化。最后，图 5.35(d)显示了几何校正过的图像。这幅图像与原始图像的质量相同。

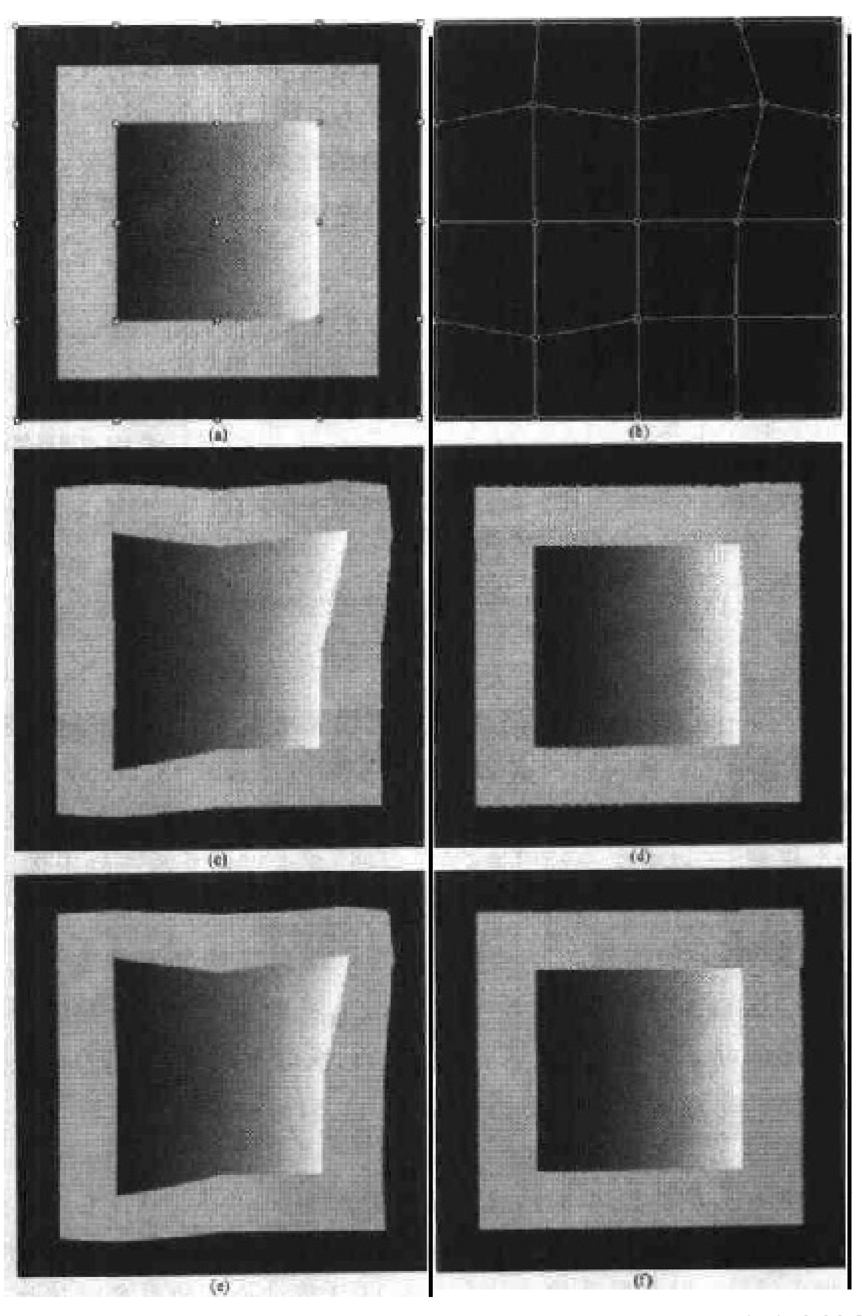


图 5.34 (a)显示连接点的图像，(b)几何失真后的连接点，(c)用最近邻点内插失真的图像，(d)复原结果，(e)使用双线性内插的失真图像，(f)复原图像

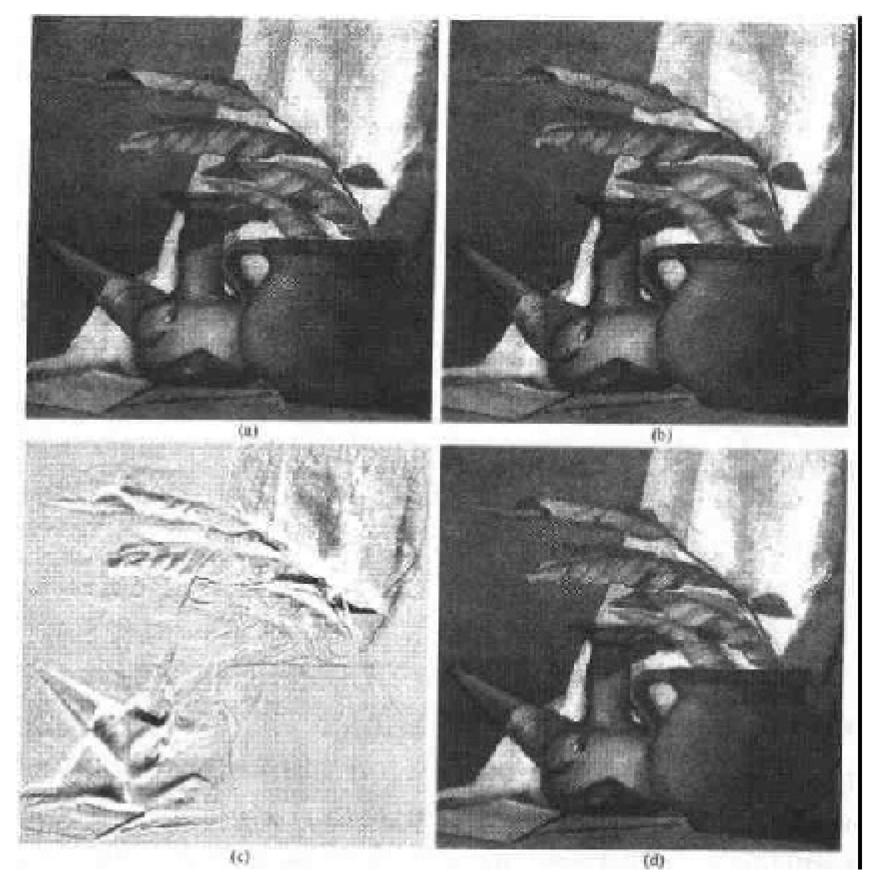


图 5.35 (a)几何失真前的图像,(b)用与图 5.34 (e) 相同的参数几何失真的图像,(c)(a)和(b)间的差,(d)几何复原的图像

小结

这一章的主要结果建立在假定图像退化为线性的位置不变过程并带有加性噪声的模型,噪声与图像值无关。甚至当这些假设并不完全符合时,用前几章开发的方法得到有用的结果也常常是可能的。

在这一章推导的复原技术基于各种最佳准则。在此文中"最佳"一词的使用涉及到严格的数学概念,而不是人类视觉系统的最佳响应。事实上,对于视觉感知知识的缺乏妨碍了图像复原问题的一般表达,图像复原问题要考虑观察者的爱好与能力。在这些限制条件下,本章介绍的概念的优势是基础方法的发展,它有合理的预测能力和坚实的基础知识的支持。

第 3 节和第 4 节中的一般复原任务(比如随机噪声减少)是在空间域使用卷积模板来执行的。我们也发现频域对减少周期性噪声和对某些重要的退化建模(如在图像获取时由运动导致的模糊)也是很理想的。并且,对于表达复原滤波器,如维纳和约束最小二乘方滤波器,频域也是很有用的工具。

第 4 节提到过,频域对实验提供了一个直觉、可靠的基础。一旦对于给定的应用令人满意的一种滤波方法被发现,其实现通常是通过设计一个数字滤波器在频域近似解决来执行的,在计算机或专用的硬件/固件系统中运行速度要快得多。数字滤波器的设计超出了本书的范围,

但是这一课题的很多参考书都包括在下面的部分中。

参考资料

在5.1节提供的线性退化模型方面的其他读物可见Castleman[1996]和Pratt[1991]的著述。Peebles[1993]的论著提供了噪声概率密度函数及其特性的一般水平的概括(5.2节)。由Papoulis[1991]所写的书更深入、更详细地概括了这些概念。5.3节的参考材料有Umbaugh[1998],Boie和Cox[1992]的书籍,Hwang和Haddad[1995],Wilburn[1998],Eng和Ma[2001]的书籍。自适应滤波器设计的一般领域对于5.3节中所讨论的自适应滤波器是一个很好的背景。由Haykin[1996]所著的书籍对这一题目进行了很好的介绍。5.4节介绍的滤波器是第4章材料的直接延伸。对于5.5节材料的其他读物可见Rosenfeld和Kak[1982]及Pratt[1991]的著述。

估计退化函数的课题(5.6节)是当前一个有趣的领域。估计退化函数的早期技术由Andrews和Hunt[1977],Rosenfeld和Kak[1982],Bates和McDonnell[1986]及Stark[1987]给出。因为退化函数很少能精确知道,过去一些年有大量的技术被提出,其中,强调了特殊复原。例如,Geman和Reynolds[1992],Hurn和Jennison[1996]提出保护灰度尖锐过渡以加强尖锐化,Boyd和Meloche[1998]集中于退化图像中的细小物体复原。处理图像模糊技术的例子是Yitzhaky等人[1998],Harikumar和Bresler[1999],Mesarović[2000]及Giannakis和Heath[2000]的著述。序列图像的复原也是考虑的重点。由Kokaram[1998]所著的书籍提供了这一领域很好的基础。

对于5.7节到5.10节讨论的滤波方法,这些年在图像处理方面的书籍和论文中已用各种方法进行了解释。有两个主要方法作为这些滤波器的开发基础。一个是使用矩阵理论基于通常的表示法,如Andrews和Hunt[1977]的介绍。这个方法精致而通用,但是对于新涉猎该领域的人来说比较困难,因为它缺乏直观性。直接基于频率域的方法(在这一章所遵循的方法)对于初次接触复原的人来说通常比较容易,但是缺少矩阵方法统一的数学严谨性。两种方法可达到相同的结果,但是根据我们教授这些材料的经验,初涉这一领域的学生更喜欢后一种方法。对于5.7节到5.10节讲解的滤波概念的补充读物是Castleman[1996],Umbaugh[1998],Petrou和Bosdogianni[1999]的著述。最后的参考书在二维频率域滤波器和相应的数字滤波器之间提供了很好的联系,二维数字滤波器的设计见Lu和Antoniou[1992]的著述。虽然在本章没有包括其内容,但计算机化的轴向形态学领域有时作为复原的课题也被包括进来。对这一领域,Kak和Slaney[2001]给出了全面的介绍。5.11节进一步的基本读物见Sonka等人[1999]的著述。Unser等人[1995]和Carey等人[1999]的论文也很重要。

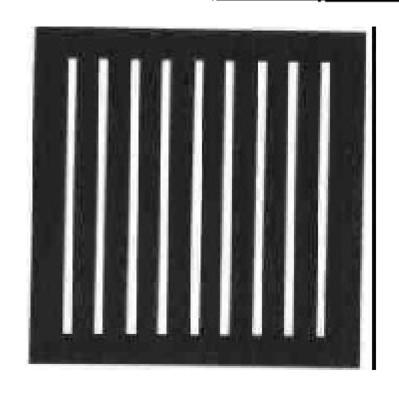
习题

★5.1 图中的白条是7像素宽,210像素高。两个白条之间的宽度是17像素,当应用下面的处理时图的变化结果是什么?

(a) 3×3算术均值滤波。

(b) 7×7算术均值滤波。

(c) 9×9算术均值滤波。



注意：这个问题及跟随它的一个关系到滤波图像的问题看来有些麻烦。然而，它们值得尝试，因为它们可帮助理解这些滤波器如何工作。理解了特殊图像滤波效果之后，回答可以是简短的关于结果图像的描述。例如，"结果图像将由3个像素宽，206个像素高的白条组成"。一定要描述白条的任何变形，例如，圆角。可以忽略边界影响，其中模板只部分地包含图像像素。

5.2 用几何均值滤波重复习题5.1的问题。

★**5.3** 用谐波均值滤波重复习题5.1的问题。

5.4 用逆谐波均值滤波($Q=1$)重复习题5.1的问题。

★**5.5** 用逆谐波均值滤波($Q=-1$)重复习题5.1的问题。

5.6 用中值滤波重复习题5.1的问题。

★**5.7** 用最大值滤波重复习题5.1的问题。

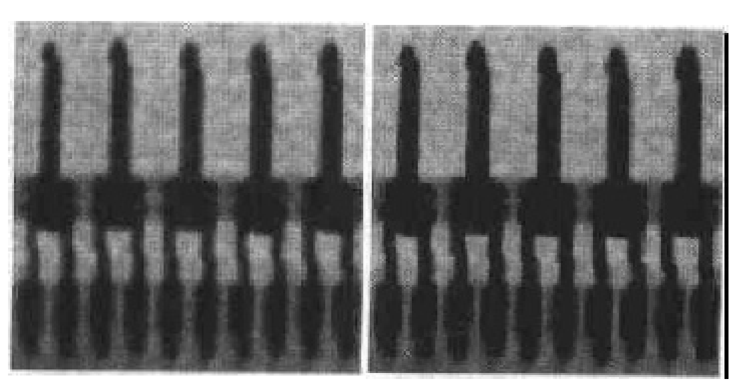
5.8 用最小值滤波重复习题5.1的问题。

★**5.9** 用中点滤波重复习题5.1的问题。

5.10 所示的是从顶部提取的两个子图像，分别是图5.7(c)和(d)的右上角图像，左边的子图像是使用尺寸为3×3的算术均值滤波器的结果，另一个使用了同样尺寸的几何均值滤波器。

★(a) 解释为什么用几何均值滤波的子图像有一点模糊。提示：用一维阶跃边缘剖面图分析(见图3.38的例子)。

(b) 解释为什么右边图像的黑色部分更粗了？



5.11 根据式(5.3.6)给出的逆谐波滤波回答下列问题：

(a) 解释为什么当Q是正值的时候滤波对去除"胡椒"噪声有效？

(b) 解释为什么当Q是负值的时候滤波对去除"盐"噪声有效？

(c) 解释为什么当选择的 Q 值不当时滤波给出较差的结果?

(d) 讨论当 $Q=-1$ 时滤波的特性。

(e) 讨论在恒定灰度级区域滤波的特性(包括 Q 正负的情况)。

★**5.12** 获得对应于式(5.4.3)到式(5.4.1)的带阻滤波器的带通滤波器的等式。

5.13 求出对应于在5.4.3节中讨论的陷波带阻滤波器的陷波带通滤波器的等式,说明当 $u_0=v_0=0$ 时它们变成低通滤波器。

★**5.14** 说明二维正弦函数的傅里叶变换

$$f(x,y)=A\sin(u_0x+v_0y)$$

是共轭脉冲对:

$$F(u,v)=-\mathrm{j}\frac{A}{2}\left[\delta\left(u-\frac{u_0}{2\pi},v-\frac{v_0}{2\pi}\right)-\delta\left(u+\frac{u_0}{2\pi},v+\frac{v_0}{2\pi}\right)\right]$$

提示:用式(4.2.3)的连续傅里叶变换并以指数项描述正弦。

5.15 用式(5.4.19)推导式(5.4.21)。

★**5.16** 设有一个线性位置不变的图像退化系统具有如下的冲激响应:

$$h(x-\alpha,y-\beta)=\mathrm{e}^{-[(x-\alpha)^2+(y-\beta)^2]}$$

假定输入到系统的是无穷小宽度的线组成的图像,位于 $x=a$,并由 $f(x,y)=\delta(x-a)$ 模型化,δ 是一个冲激函数,假设无噪声,输出图像 $g(x,y)$ 是什么?

5.17 在图像获取中,一幅图像进行了在垂直方向上的均匀线性运动,用时 T_1。运动方向然后转为水平,时间间隔 T_2,假设图像改变方向所用时间可以忽略,同样快门开关时间也可以忽略,给一个表达式描述模糊函数 $H(u,v)$。

★**5.18** 考虑在 x 方向均匀加速导致的图像模糊问题。如果图像在 $t=0$ 静止,并用均匀加速 $x_0(t)=at^2/2$ 加速,对于时间 T,找出模糊函数 $H(u,v)$,可以假设快门开关时间忽略不计。

5.19 一个空间探测器被设计用于从行星发送图像,探讨着陆方法。在着陆的最后阶段一个控制推进器坏了,导致飞行器沿垂直轴急速旋转。在着陆期间的最后两秒钟发送的图像被模糊为其圆周运动的序列。摄像机位于探测器的底部,沿着垂直轴,并指向下方。遗憾的是,飞行器的转动也是沿着垂直轴,因此,图像由均匀旋转运动模糊了。在每一幅图像获取期间飞行器旋转 $\pi/8$ 弧度。假设在图像获取期间垂直运动可以忽略,对复原图像给出解。

★**5.20** 所示图像是一个模糊的心脏的二维再现立体投影。已知每个图像右边底部的十字线是3像素宽,30像素长,在模糊之前有255的灰度值,请提供一个过程,指出怎样应用刚才提供的信息得到模糊函数 $H(u,v)$。



(原图由 GE 医学系统提供)

5.21 某一个 x 光成像几何产生了模糊退化,它可用空间圆对称函数检测图像的卷积来模型化:

$$h(r)=[(r^2-\sigma^2)/\sigma^4]\mathrm{e}^{-r^2/2\sigma^2}$$

在这里 $r^2=x^2+y^2$。说明频率域退化由下面表达

式给出：

$$H(u,v) = -\sqrt{2\pi}\sigma(u^2+v^2)e^{-2\pi^2\sigma^2(u^2+v^2)}$$

★**5.22** 使用习题5.21中的传递函数，对维纳滤波器给出描述，假定噪声功率谱和未退化信号之比为常数。

5.23 使用习题5.21中的传递函数，对约束最小二乘方滤波器给出结果描述。

5.24 假定图5.1的模型是线性且位置不变的，并示出输出功率谱由下式给出：

$$|G(u,v)|^2 = |H(u,v)|^2|F(u,v)|^2 + |N(u,v)|^2$$

参阅式(5.5.17)和式(4.2.20)。

5.25 Cannon[1974]建议过一个复原滤波器 $R(u,v)$，满足条件：

$$|\hat{F}(u,v)|^2 = |R(u,v)|^2|G(u,v)|^2$$

并且以强制复原图像的功率谱为前提，$|\hat{F}(u,v)|^2$ 等于原始图像的功率谱 $|F(u,v)|^2$。

★(**a**) 在 $|F(u,v)|^2$，$|H(u,v)|^2$，$|N(u,v)|^2$ 中找出 $R(u,v)$，提示：参阅图5.1、式(5.5.17)和习题5.24。

(**b**) 用在(a)中的结果，以式(5.8.2)的形式说明这一结果。

5.26 一位天文学家用大倍数望远镜观察他的图像有点模糊。望远镜镜头聚焦图像于高分辨率CCD成像阵列，图像然后由望远镜电子学方法转换为数字图像。试图用传导控制实验室的实验改善这种情况，由于望远镜元件的尺寸和重量，用镜头和传感器是不可能的。天文学家听说你是成功的图像处理专家，要求你帮助他为使图像稍微尖锐一些提出一个数字图像处理解决方案。将如何解决这个问题(仅能得到天体图像)。

★**5.27** 一位考古学教授在做流通货币方面的研究。在罗马帝国时期，最近认识到4个罗马硬币对他的研究是决定性的，它们被列在伦敦大英博物馆的馆藏目录中，遗憾的是，他到达那里后，被告知现在硬币已被盗了，进一步在这部分进行研究表明，依赖博物馆保存的一些照片是可靠的。遗憾的是，硬币的照片模糊了，数据上的点和小的标记不能读出。模糊的原因是摄取照片时照相机散焦。作为一名图像处理专家和教授的朋友，要求你帮助决定是否计算机处理能被用于复原图像，指导教授在哪里能读这个标记。且用于拍摄该图像的原照相机一直能用，像同一时期其他有代表性的硬币那样。提出解决这一问题的过程。

5.28 不使用四边形，而请用5.11节的三角区域去完成空间变换和灰度插补。对三角区域将用什么公式模拟式(5.11.5)、式(5.11.6)和式(5.11.7)?

第 6 章　彩色图像处理

在图像处理中运用颜色受两个主要因素推动。第一,颜色是一个强有力的描绘子,它常常可简化目标物的区分及从场景中抽取目标;第二,人可以辨别几千种颜色色调和亮度,相形之下只能辨别几十种灰度层次。第二个因素对于人工图像分析特别重要。

彩色图像处理可分为两个主要领域:全彩色处理和伪彩色处理。在第一类中,图像用全彩色传感器获取,如彩色电视摄像机或彩色扫描仪。在第二类中,对特定的单一亮度或亮度范围赋予一种颜色。到目前为止,数字彩色图像处理多为伪彩色处理。

然而,在过去的十年,处理彩色图像的彩色传感器和硬件的价格变得更容易接受,结果导致全彩色图像处理技术的应用日益广泛,包括印刷、可视化和互联网应用。在后续的讨论中,前而各章涉及的灰度级方法可直接用于彩色图像处理,这一点是很明显的。其余的将按照本章推导的彩色空间特性加以重新阐述。本书讨论的这些技术还远没有穷尽,只是涉及了彩色图像处理的常用方法。

6.1　彩色基础

虽然人的大脑感知和理解颜色所遵循的过程是一种生理心理现象,这一现象还远没有完全了解,但颜色的物理性质可以由实验和理论结果支持的基本形式来表示。在 1666 年,艾萨克·牛顿(Isaac. Newton)发现了一个现象,当一束太阳光通过一个玻璃棱镜时,出现的光束不是白的,而是由从一端为紫色到另一端为红色的连续彩色谱组成。如图 6.1 所示,彩色谱可分为 6 个宽的区域:紫色、蓝色、绿色、黄色、橘红色和红色。当观察全彩色时(图 6.2),色谱末尾的颜色不是突变的,而是一种颜色连续地平滑过渡到下一种颜色。

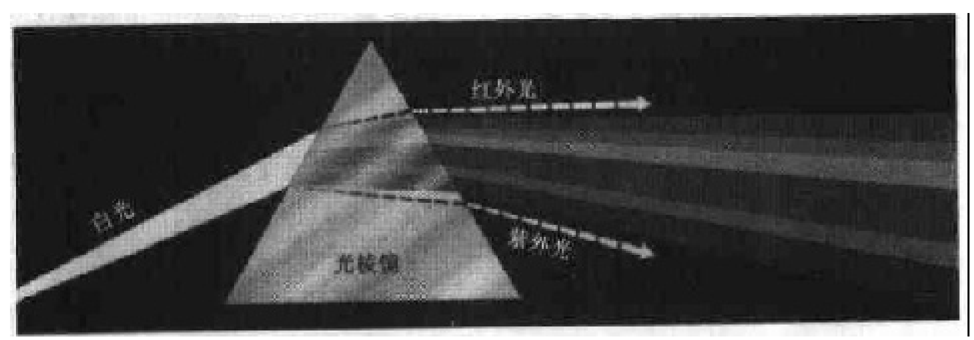


图 6.1　白光通过棱镜时看到的彩色谱(由 GE 公司 Lamp Business 分公司提供)

基本上,人类和某些其他动物接收的一个物体的颜色由物体反射光的性质决定。正如图 6.2所示,可见光是由电磁波谱中相对较窄的波段组成的。一个物体反射的光如果在所有可见光波长范围内是平衡的,对观察者来说显示白色。然而,若一个物体对有限的可见光谱范围反射,则物体呈现某种颜色。例如,绿色物体反射具有 500 ~ 570 nm 范围的光,吸收其他波长光的多数能量。

光特性是颜色科学的核心。如果光是消色的(缺乏颜色),它的属性仅仅是亮度或者数值,消色光就是观察者看到的黑白电视的光,就是迄今为止我们讨论过的图像处理的固有分量。正如在第 2 章定义并用大量篇幅讨论的那样,灰度仅提供了一个度量亮度的标量,它的范围从黑到灰最后到白。

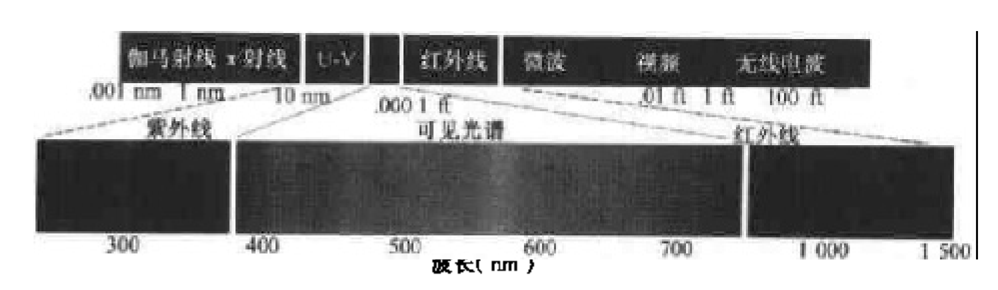


图 6.2　可见范围电磁波谱的波长组成(由 GE 公司 Lamp Business 分公司提供)

彩色光大约覆盖电磁波谱 400 ~ 700 nm 的范围,3 个基本量用来描述彩色光源的质量:辐射率,光强和亮度。辐射率是从光源流出能量的总量,通常用瓦特(W)度量。光强用流明度量,它给出了观察者从光源接收的能量总和的度量。例如:光从工作在远红外波谱范围的光源中发出,它可能具有实际意义上的能量,但是观察者却很难感觉到;它的光强几乎是零。最后,亮度是一个主观描绘子,它实际上是不可能度量的。它具体化了彩色强度的概念并是描述彩色感觉的一个关键参数。

正如 2.1.1 节提到的,人眼的锥状细胞是负责彩色视觉的传感器,详细的实验结果已确定,在人眼中的 6 ~ 7 百万锥状细胞中可分为 3 个主要的感觉类别,它们对应红、绿和蓝。大约 65% 的锥状细胞对红光敏感,33% 对绿光敏感,只有 2% 对蓝光敏感(但是蓝锥状细胞对蓝光更敏感)。图 6.3 显示了眼睛对红、绿、蓝光吸收的平均实验曲线。由于人眼的这些吸收特性,被看到的彩色是所谓的原色红(R)、绿(G)、蓝(B)的各种组合。为标准化起见,CIE(国际照明委员会)在 1931 年设计了下面的特定波长值为主原色:蓝 = 435.8 nm,绿 = 546.1 nm,红 = 700 nm。这一标准在图 6.3 所示的详细实验曲线于 1965 年采用之前就做了规定。这样,CIE 标准只是实验数据的近似。从图 6.2 和图 6.3 注意到,没有单一颜色可称为红、绿、蓝。还有,应记住,为标准化目的而定的特定三原色波长并不意味着 RGB 分量单独作用就能产生所有谱色,这一点很重要。原色这一词的使用已被广泛地误解为:3 个标准原色当以各种强度比混合在一起时可以产生所有可见彩色。正如很快会看到的那样,这一解释是不正确的,除非波长允许变化,在这种情况下,将不再有固定的标准原色了。

原色相加可产生二次色,如深红色(红加蓝),青色(绿加蓝)和黄(红加绿)。以正确的亮度把三原色或者一种二次色与其相反的原色相混合可产生白光。这一结果示于图 6.4(a),它还说明三原色和它们相混合产生二次色。

光原色与颜料或着色剂原色之间的区别是很重要的。后者,定义一种原色为减去或吸收光的一种原色并反射或传输另两种原色。因此,颜料的原色是深红色,青色和黄色,而二次色是红、绿和蓝。这些颜色示于图 6.4(b)。这三种颜料原色或者其相对应的补色进行合适的组合产生黑色。彩色电视接收机是色光相加的例子,许多彩色电视显像管内部是由电敏荧光粉三角形点阵形式组成的。当激发时,三色组中的每一点能产生三原色中的一种光。发射红光的荧光粉的像点亮度由显像管内的电子枪调制,该电子枪产生的脉冲与电视摄像机摄取的"红

能量"相对应。三像素组中的绿点和蓝点荧光物以相同的方式被调制。在电视接收机上观察到的效果是,三像素组每一荧光点原色加在一起并由眼睛对颜色敏感的锥状体以全彩色图像的方式接收。以这三种颜色方式每秒钟连续改变 30 幅画面,在荧光屏上可完成连续图像显示。

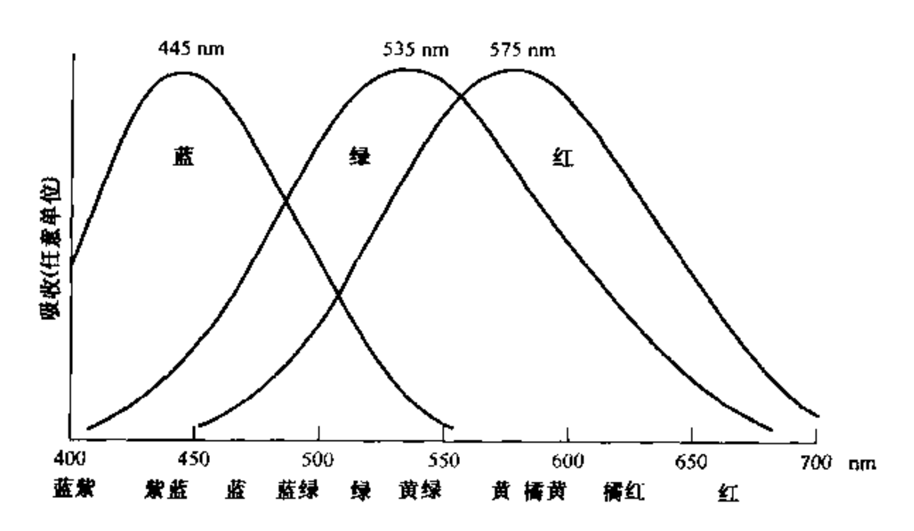


图 6.3 人眼中红、绿、蓝锥状体的波长吸收函数

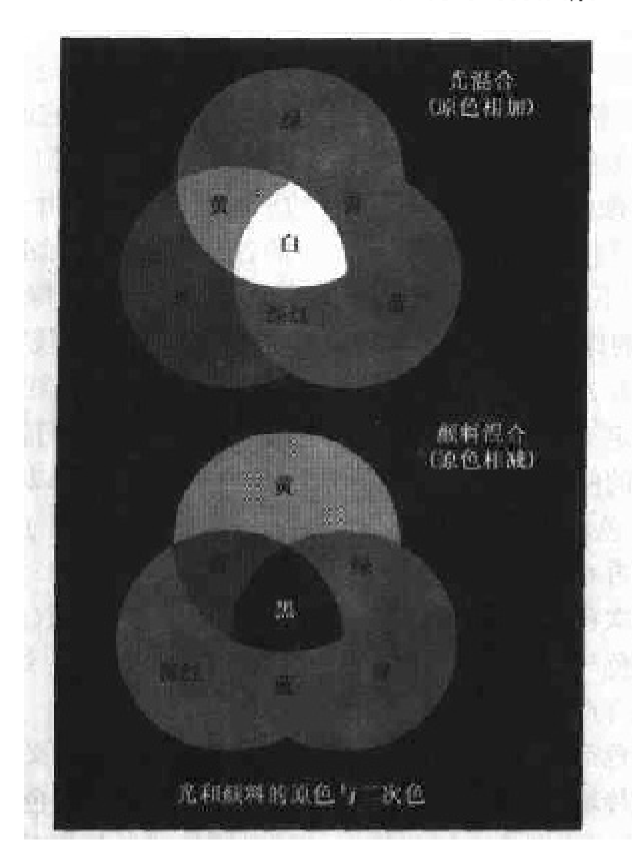


图 6.4 光及颜料的原色和二次色(由 GE 公司 Lamp Business 分公司提供)

通常用以区别颜色的特性是亮度、色调和色饱和度。正如本节前面指出的那样，亮度是色彩明亮度的概念，色调是光波混合中与主波长有关的属性，色调表示观察者接收的主要颜色。这样，当我们说一个物体是红色、橘黄色、黄色时，是指它的色调。纯谱色是全饱和的。像粉红色（红加白）和淡紫色（紫加白）是欠饱和的，饱和度与所加白光数量成反比。

色调与饱和度一起称为彩色，因此，颜色用亮度和彩色表征。形成任何特殊颜色需要的红、绿、蓝的量称做三色值，并分别表示为 X，Y 和 Z。一种颜色由三色值系数定义为：

$$x = \frac{X}{X + Y + Z} \tag{6.1.1}$$

$$y = \frac{Y}{X + Y + Z} \tag{6.1.2}$$

$$z = \frac{Z}{X + Y + Z} \tag{6.1.3}$$

从以上公式可得：[①]

$$x + y + z = 1 \tag{6.1.4}$$

对任何可见光谱内的光波长，产生与波长相对应颜色的三色值可直接从大量实验结果编制的曲线或列表中得到（Poynton[1996]）。这也可参看 Walsh[1958] 和 Kiver[1965] 的早期文献。

确定颜色的另一种方法是用 CIE 色度图。该图以 x（红）和 y（绿）函数表示颜色组成。对于 x 和 y 的任何值，其相应的（蓝）值可从式(6.1.4)得到。注意 $z = 1 - (x + y)$，例如，图 6.5 中标记为绿的点有 62% 的绿和 25% 的红成分，从式(6.1.4)得到蓝的成分约为 13%。

从380 nm的紫色到780 nm的红色的各种谱色的位置标在舌形色度图周围的边界上。这些都是示于图 6.2 的谱图中的纯色。任何不在边界上而在色度图内部的点都表示谱色的混合色。示于图6.5的等能量点与三原色百分率相对应，它表示相对于白光的 CIE标准。位于色度图边界上的任何点都是全饱和的。如果一点离开边界并接近等能量点，就在颜色中加入了更多的白光，该颜色就变成欠饱和的了。等能量点的饱和度为零。

色度图对彩色混合非常有用，因为在色度图中用连接任意两点的直线段定义所有不同颜色的变化，这些颜色可以由这两类颜色相加得到。例如，考虑图 6.5 中示出的从红点到绿点画的一条直线，如果有比绿光多的红光，则确切地表示新颜色的点将处在线段上，但与绿点相比更接近红点。类似地，从等能量点到位于色度图边界上的任一点画一线段将定义特定谱色的所有色调。

把这一过程扩展到三种颜色是很简单的。为了从色度图中任何给定的三种颜色决定得到的颜色范围，简单地对三色点的每一个画一条连线。其结果是一个三角形，三角形内的任何颜色都可以由三种原色的不同混合产生。

以任意确定颜色为顶点的三角形不能包围图 6.5 中的所有颜色范围。这个结果从图解上支持早期得出的用 3 个单一的、确定的原色不能得到所有颜色的论点。图 6.6 的三角形显示了由 RGB 监视器产生的典型的颜色范围（称做彩色全域）。在三角形内不规则的区域表示高

① 在这段文字中的 x, y, z 遵循国际惯例，这将不会与本书其他章节定义的空间坐标(x, y)相混淆。

质量的彩色打印设备的彩色域。彩色打印域的边界是不规则的,因为彩色打印是相加和相减彩色混合的结合,相比在监视器上显示颜色,控制过程要困难得多,监视器是基于三个高质量的可见光原色的相加。

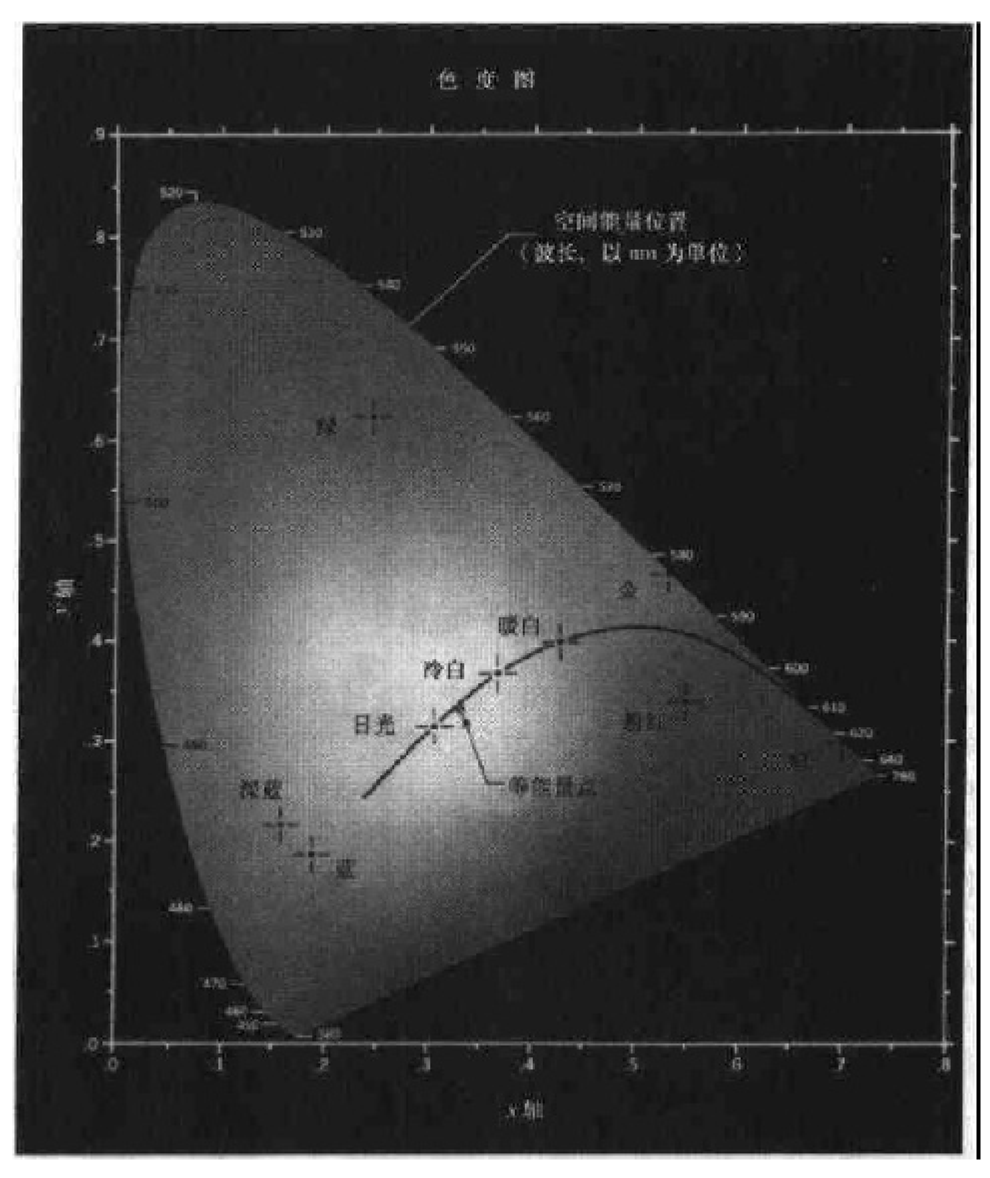


图 6.5　色度图(由 GE 公司 Lamp Business 分公司提供)

6.2 彩色模型

彩色模型(也称彩色空间或彩色系统)的用途是在某些标准下用通常可接受的方式简化彩色规范。本质上,彩色模型是坐标系统和子空间的规范。位于系统中的每种颜色都由单个点来表示。

现在所用的大多数彩色模型都是面向硬件的(如:彩色监视器和打印机)或是面向应用的。在这里,彩色操作是一个目的(如针对动画的彩色图像的创作)。在数字图像处理中,实际中最通用的面向硬件的模型是RGB(红、绿、蓝)模型。该模型用于彩色监视器和一大类彩色视频摄

像机。CMY(青、深红、黄)、CMYK(青、深红、黄、黑)模型是针对彩色打印机的。HSI(色调、饱和度、亮度)模型更符合人描述和解释颜色的方式。HSI 模型还有一个优点,就是把图像分成彩色和灰度信息,使其更适合本书开发的许多灰度处理技术。现在使用的彩色模型还有不少,这主要是由于彩色科学是一个包括许多应用场合的很宽的领域。这里试图简单地详细研究这些模型中的几个,因为它们对于图像处理更有趣和有益。然而,根据现有的任务,在这一章讨论的模型对图像处理来说是主要的模型。有了这一章的主要资料,读者理解当今所用的其他模型将不会太困难。

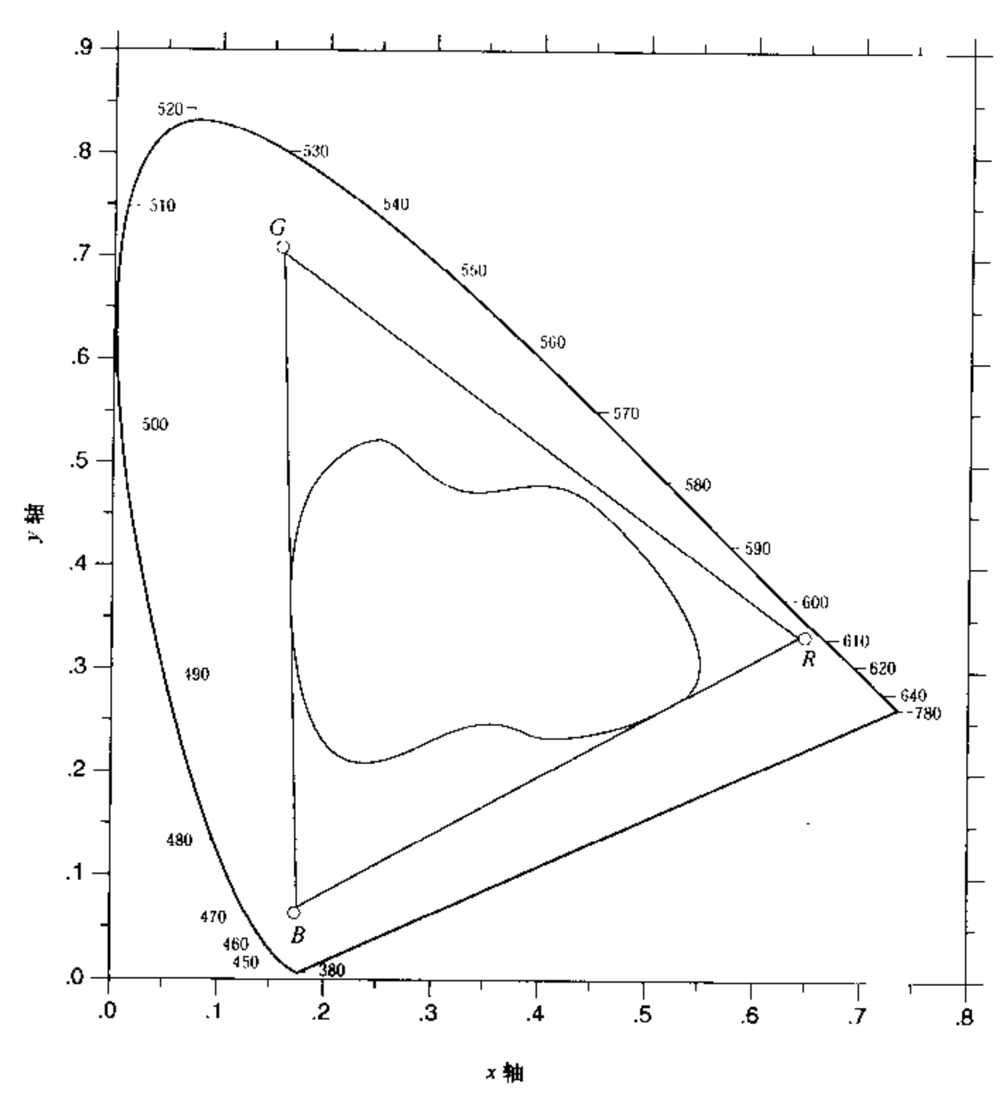


图 6.6 彩色监视器和彩色打印设备的典型彩色域

6.2.1 RGB 彩色模型

在 RGB 模型中,每种颜色出现在红、绿、蓝的原色光谱分量中,这个模型基于笛卡儿坐标系统,所考虑的彩色子空间是图 6.7 所示的立方体。图中,R, G, B 位于 3 个角上;青、深红和黄位于另外 3 个角上,黑色在原点处,白色位于离原点最远的角上。在该模型中,灰度等级沿着这两点的连线分布。在本模型中,不同的颜色处在立方体上或其内部,并可用从原点分布的

向量来定义。为方便起见,假定所有的颜色值都归一化,则图 6.7 所示的立方体就是一个单位立方体,即所有 R,G,B 的值都在[0,1]范围内取值。

在 RGB 彩色模型中,所表示的图像由 3 个图像分量组成,每一个分量图像都是其原色图像。当送入RGB监视器时,这三幅图像在荧光屏上混合产生一幅合成的彩色图像。在RGB空间,用以表示每一像素的比特数叫做像素深度。考虑RGB图像,其中每一幅红、绿、蓝图像都是一幅 8 比特图像,在这种条件下,每一个RGB彩色像素[(R,G,B)值 3 个一组]称为有 24 比特深度(3 个图像平面乘上每平面比特数)。全彩色图像常用来定义 24 比特的彩色图像。在 24 比特RGB图像中颜色总数是$(2^8)^3$ = 16 777 216。图 6.8 显示了与图 6.7 相对应的 24 比特彩色立方体。

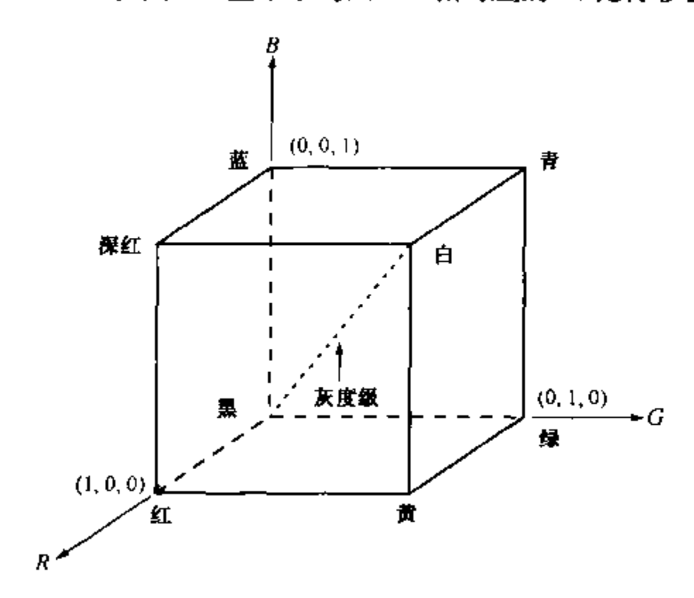


图 6.7 RGB 彩色立方体示意图,沿着主对角线从原点的黑色到点(1,1,1)的白色是灰度值

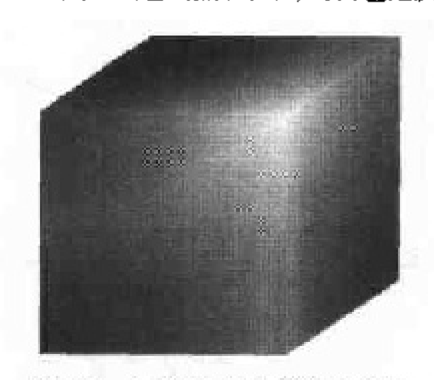


图 6.8 RGB 24 比特彩色立方体

例 6.1 产生隐藏面和 RGB 彩色立方体的横截面

图 6.8 所示的立方体是实心的,在前边的章节提到过立方体由$(2^8)^3$ = 16 777 216种颜色组成。观察这些颜色的一个方便的方法是产生一个彩色平面(立方体的表面或横切面)。彩色平面可通过固定三种颜色中的一个而允许其他两个变化的方法完成。例如,在图 6.7 和图 6.8 中,通过立方体中心并与 GB 平面平行的横截面是$(127,G,B)$平面,其中 $G,B=0,1,2,\cdots,255$。在这里,用的是实际像素值,而不是为方便在[0,1]范围内的数学归一化值。因为前者是在计算机中实际采用来产生颜色的值。图 6.9(a)显示了一幅横截面图像,该图像是把三幅独立的分量图

像送入彩色监视器上观察到的。在分量图像中 0 代表黑色,255 代表白色(注意这些是灰度图像)。最后,图 6.9(b)显示了在图 6.8 中用同样方法产生的立方体的三幅隐藏平面。

值得注意的是获得彩色图像基本是图 6.9 所示的相反过程。一幅彩色图像可用三个分别对红、绿、蓝敏感的滤色片得到。当我们用装有一种滤色镜的单色摄像机观看彩色场景时,结果是一幅单色图像,其亮度与滤色片的响应成正比。用每一个滤色片重复这一过程产生三幅单色图像,这些图像就是彩色场景的 RGB 分量图像(实际上,通常彩色图像传感器将这一过程集成在一个装置中)。从图 6.9(a)所示的形式显示这三幅 RGB 分量图像就会产生一幅原彩色场景的 RGB 彩色复现图像。

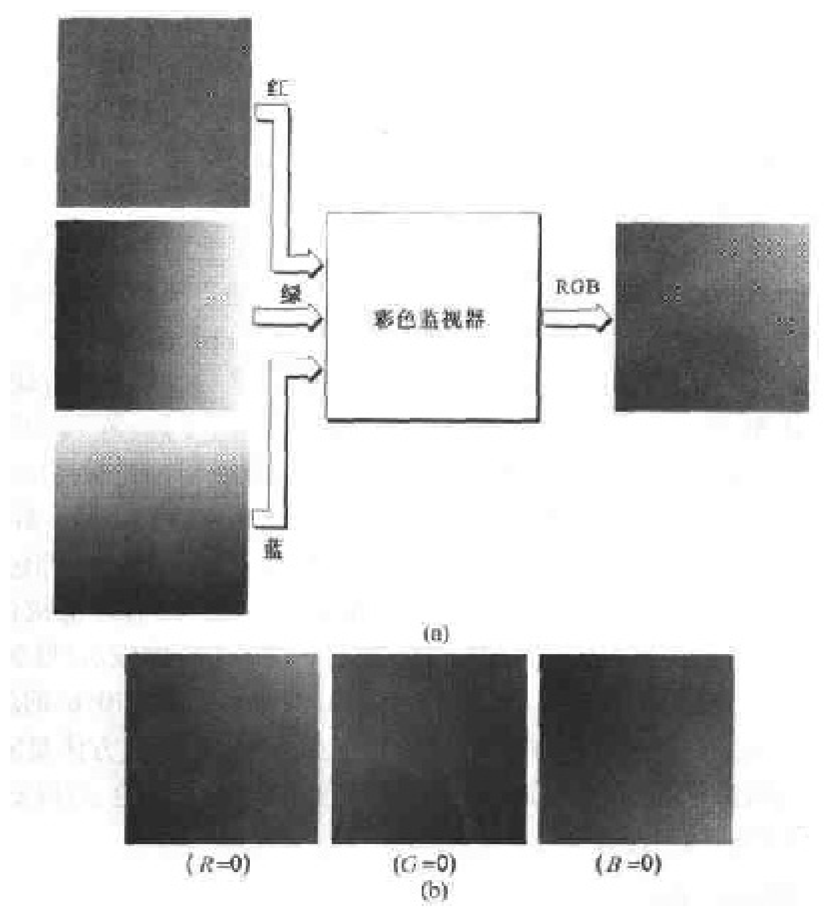


图 6.9　(a)产生彩色平面横截面(127, G, B)的 RGB 图像,(b)图6.8的彩色立方体的三幅隐藏平面

虽然高端显示卡和监视器提供 24 比特 RGB 图像的可接受重现,但今天所用的许多系统还限制在 256 种颜色。另外,有时不使用几百种颜色,而是使用几种颜色的情况也有大量的应用。在 6.3 节讨论的伪彩色图像处理技术就是一个较好的例子。在应用的多种系统中一种彩色子系统值得考虑,其合理的与观察者无关的硬件性能可保证彩色的真实再现。这种彩色子系统称为安全 RGB 彩色子集,或者全系统安全彩色集。在网络应用中叫做无损 Web 彩色,或者无损监视彩色。

假定 256 种颜色是最小颜色数,这些颜色可用任何系统真实再现。该系统可以显示所希望的彩色,用可接受的标准符号来表示这些颜色是很有用的。已知 256 种颜色中的 40 种用各种操

作系统进行不同的处理,仅留下216种颜色是各种系统通用的。这216种颜色成为事实上的安全彩色,特别是在网络应用中更是如此。无论在何时应用,希望大多数人观察到的颜色都一样。

216种安全彩色的每一种如以前一样都可以由RGB值形成,但每一个值仅可能取0,51,102,153,204或255。这样,RGB三元组数值给出$(6)^3=216$可能值(注意所有值都可被3除)。通常这些值可用16进制数系统来表示,如表6.1所示。

表6.1 安全色中RGB分量的可用值

数制系统	彩色值					
16进制	00	33	66	99	CC	FF
10进制	0	51	102	153	204	255

16进制数0,1,2,…,9,A,B,C,D,E,F对应10进制数0,1,2,…,9,10,11,12,13,14,15。还可以表示为$(0)_{16}=(0000)_2$和$(F)_{16}=(1111)_2$。例如,$(FF)_{16}=(255)_{10}=(11111111)_2$并且两个16进制数组构成8比特的一个字节。

因为取三个数形成RGB彩色,每个安全色由表6.1中的三个两位16进制数形成。例如,最纯净的红色是FF0000。000000和FFFFFF分别表示黑色和白色。类似的结果可用更熟悉的十进制表示得到。例如,最亮的红色在十进制表示中为$R=255$(FF)和$G=B=0$。

图6.10(a)显示了由RGB转换而来的216种安全颜色。第一行左上角方块数值是FFFFFF(白),它右边的方块数值为FFFFCC,第三块为FFFF99,等等。同一阵列的第二行其值是FFCCFF,FFCCCC,FFCC99,等等。阵列中最后一个方块为FF0000(可能是最亮的红色)。第二阵列的右边正好由CCCCCC开始,并以相同的方式前进,剩下的4个阵列也是一样。最后一个阵列的最后一个方块(底部右方)为000000(黑)。注意,不是所有8比特灰色都包含在216种安全色中,这一点很重要。图6.10(b)显示了在256色RGB系统中对所有可能灰色的16进制编码。这些值的一部分在安全色集的外边,但在多数显示系统中可正常表示(根据它们的相对强度)。来自安全彩色组的灰度$(KKKKKK)_{16}$,$K=0,3,6,9,C,F$,示于图6.10(b)的最下边。

图6.11显示了RGB安全彩色立方体。不像图6.8所示的全色立方体是实心的,图6.11的立方体仅在表面有效。正如图6.10(a)所示,每一平面都有36种颜色,所以安全彩色立方体所有表面由216种不同的颜色覆盖。

6.2.2 CMY和CMYK模型

正如6.1节指出的那样,青、深红和黄色是光的二次色,换句话说,它们是颜料的原色。例如,当青色颜料涂覆的表面用白光照射时,从该表面反射的不是红光,而是从反射的白光中减去红色,白光本身是等量的红、绿、蓝光的组合。

大多数在纸上沉积彩色颜料的设备,如彩色打印机和复印机,要求输入CMY数据或在内部做RGB到CMY的转换。这一转换就是执行一个简单操作:

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{6.2.1}$$

这里再次假设所有的彩色值都归一化为[0,1]范围。式(6.2.1)显示了从涂覆青色颜料的表面反射的光不包含红色(即公式中$C=1-R$)。与此相似,纯深红色不反射绿色,纯黄色不反射

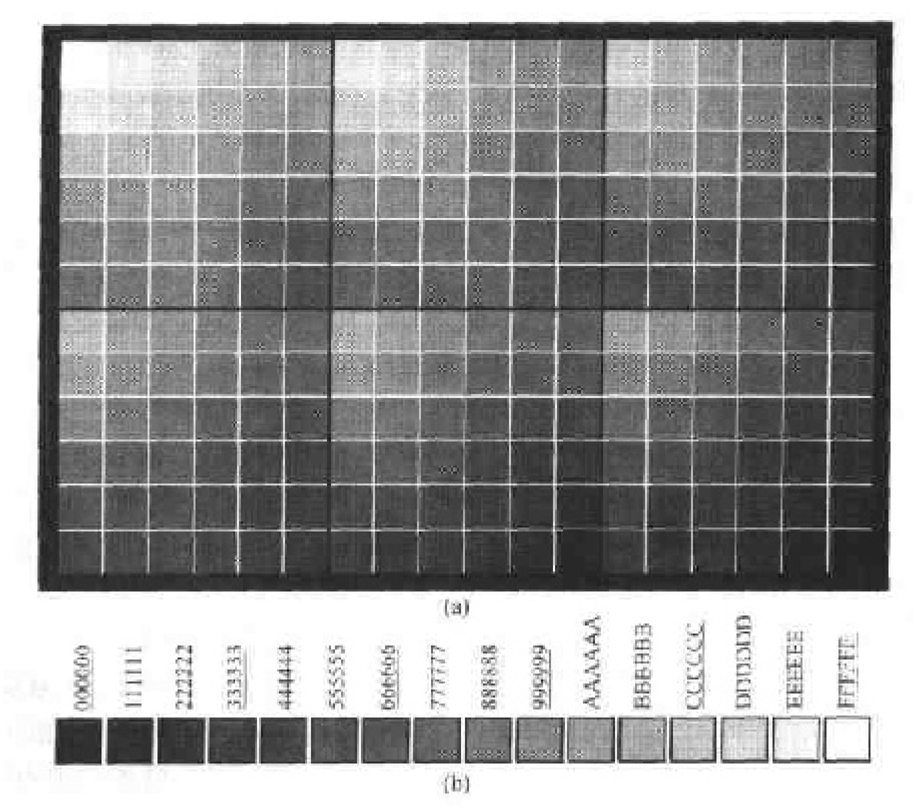


图 6.10　(a)216 种安全 RGB 色,(b)在 256 色 RGB 系统中的所有灰度(底部显示了安全彩色组的部分灰度)

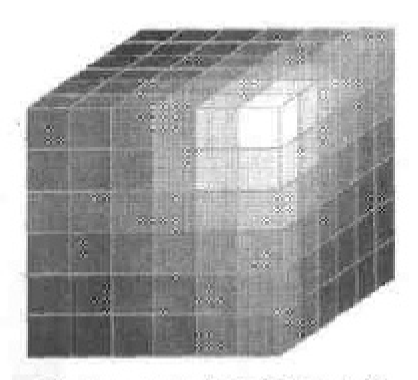


图 6.11　RGB 安全彩色立方体

蓝色。式(6.2.1)还显示RGB值可以很容易地通过 1 减去CMY值从CMY集中得到,在图像处理中,这一模型主要用于产生硬拷贝输出,因此,从CMY到RGB反向操作通常没有实际意义。

根据图 6.4 所示,等量的颜料原色(青、深红和黄色)可以产生黑色。实际上,为打印组合这些颜色产生的黑色是不纯的。因此,为了产生真正的黑色(在打印中起主要作用的颜色)加入了第四种颜色——黑色,提出了CMYK彩色模型。这样当出版商说到"四色打印"时,是指CMY彩色模型的三种原色再加上黑色。

6.2.3　HSI 彩色模型

正如我们所见到的那样,在 RGB 和 CMY 模型下产生彩色和从一种模型转换到另一种模型是一种较简单的过程。如上所述,这些彩色系统对硬件实现很理想。另外,RGB系统与人眼

很强地感觉红、绿、蓝三原色的事实能很好地匹配。遗憾的是,RGB,CMY和其他类似的彩色模型不能很好地适应实际上人解释的颜色。例如,首先,它没有涉及用组成其颜色的每一原色的百分比给出一辆汽车的颜色。此外,我们不认为彩色图像是由3幅原色图像合成一幅单一图像。

当人观察一个彩色物体时,用色调、色饱和度和亮度描述它。回顾一下6.1节的讨论,色调是描述纯色的属性(纯黄色、橘黄或红色),反过来,饱和度给出一种纯色被白光稀释的程度的度量。亮度是一个主观的描述子,实际上它是不可能测量的。它体现了无色的强度概念,并且是描述彩色感觉的关键参数。我们清楚地知道强度(灰度)是单色图像最有用的描述子。这一量是可测量的并很容易解释。即将提出的这个模型叫做HSI(色调、饱和度和强度)彩色模型,该模型可在彩色图像中从携带的彩色信息(色调和饱和度)里消去强度分量的影响,结果,HSI模型对于开发基于彩色描述的图像处理方法是一个理想的工具,这种彩色描述对人来说是自然的、直观的。然后人们便可发展和使用这些算法。可以说RGB对图像彩色的产生是理想的(用彩色摄像机捕获图像或者把图像显示在监视器屏幕上),但是对彩色描述上的应用有较多的限制。下面的资料仅提供一个非常有效的方法。

如例6.1中讨论过的,可以用3幅单色灰度图像观察RGB彩色图像(描述红、绿和蓝),因此,能够从RGB图像中提取强度并不奇怪。如果用图6.7的彩色立方体并置黑顶点为(0,0,0),白顶点为(1,1,1),如图6.12(a)所示,则这一点就更加清楚了。正如图6.7指出的强度(灰度级)沿着连接两顶点的连线分布那样,如图6.12的排列所示,连接黑和白顶点的线是垂直的。这样,如果要确定图6.12中任何点的强度分量,可以简单地通过一个垂直于强度轴的平面,平面包含该彩色点,该平面与强度轴的交点就给出了在[0,1]范围内的强度值。还注意到,彩色的饱和度(纯度)在强度轴作为距离的函数增长。事实上,强度轴上点的饱和度是零。事实证明沿着这条轴线的所有点都是灰度。

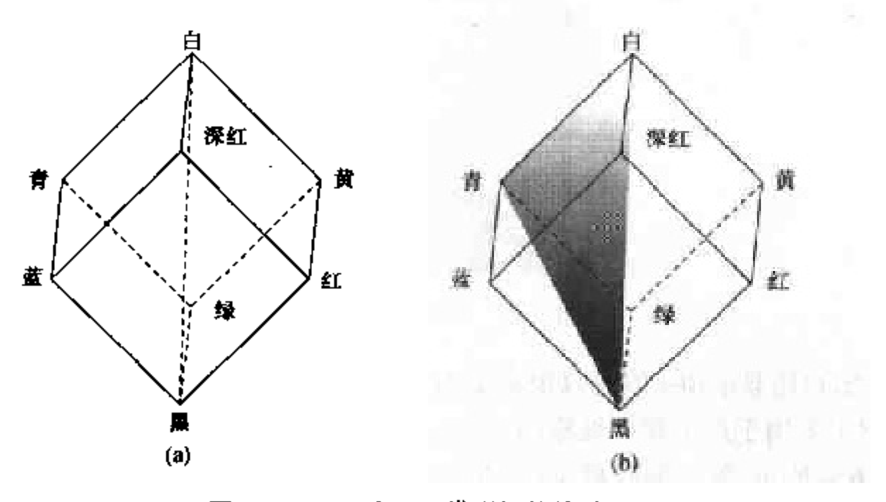


图6.12 RGB和HSI模型间的关系

为了了解如何从给定的RGB点确定色调,考虑图6.12(b),图中显示了由3个点(黑、白和青)定义的平面,黑点和白点包含在平面内的事实告诉我们强度轴也在平面内。而且,看到位于由强度轴和立方体边界决定的平面内的所有点都有相同的色调(在这种情况是青色)。回顾6.1节可得到相同的结论,即所有颜色都是由位于那些颜色定义的三角形内的三种颜色产生的。如果这些点的两点是黑和白,第三点是彩色点,在三角形上的所有点都有相同的色调,因为黑和白分量不能改变色调(当然在这个三角形中点的强度和饱和度是不同的)。沿着垂直强

度轴旋转色调平面，可得到不同的色调。从这些概念，可以得出这样的结论，HSI 空间要求的色调、饱和度和强度可以从RGB彩色立方体得到。这就是说，可以把任何RGB点转换为相应的HSI 彩色模型的点（采用描述前边讨论的推理方式的几何公式）。

关于图 6.12 立方体结构和相应HSI空间的关键一点是，垂直强度轴和位于垂直于该轴的平面的彩色点轨道表示 HSI 空间。当平面沿强度轴向上或向下移动时，由平面与立方体表面构成的横截面决定的边界不是呈三角形就是六边形。这种情况如图6.13(a)所示，在立方体沿灰度轴向下看，就会清楚得多。

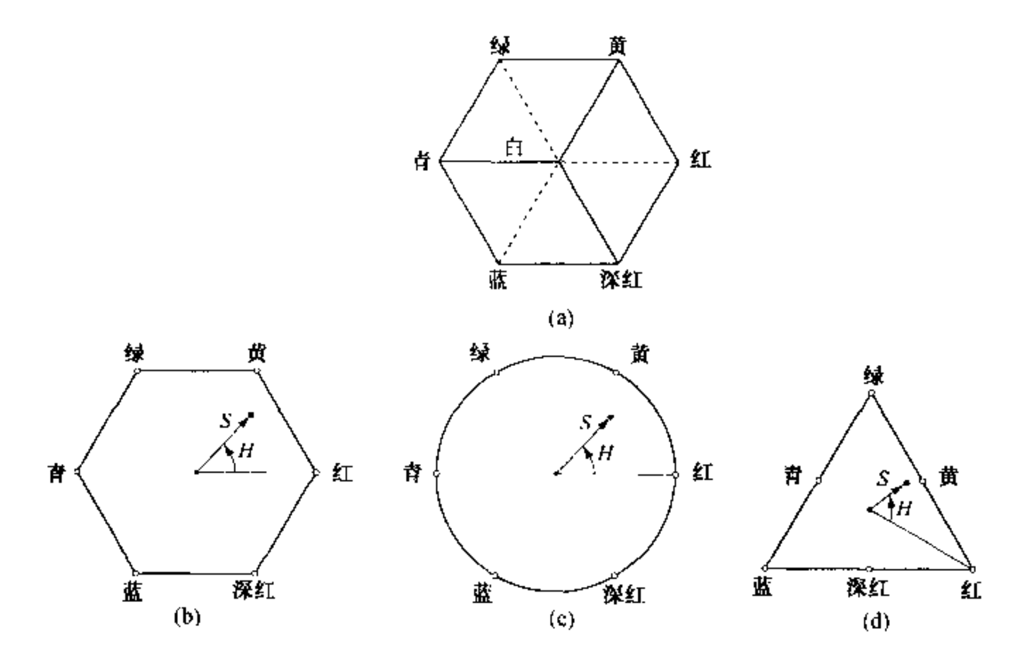


图 6.13　HSI 彩色模型中的色调和色饱和度

在这个平面中可以看到原色是按120°分隔的，二次色与原色相隔60°，这意味着二次色之间也相隔120°。图6.13(b)显示了相同的六边形和任意的一个彩色点（用一点显示）。该点的色调由来自某参考点的一个角度来决定。通常（不总是）与红轴的 0°角指定 0 色调，从这里开始色调逆时针增长。饱和度（距垂直轴的距离）是从原点到该点的向量长度。注意，原点是由彩色横截面与垂直强度轴定义的。HSI彩色空间的重要分量是垂直强度轴、到一彩色点的向量长度和这一向量与红轴的角度。因此，如图6.13(c)和(d)所示，HSI平面以刚刚讨论的六边形、三角形，甚至一个圆形的形式出现并不奇怪。实际上选择什么形状并没关系，因为这些形状中的任何一个都可以通过几何变换变为其他两种。图 6.14 显示了基于彩色三角形和圆形的HSI模型。

图中的黑点是一个任意彩色点。与红轴的夹角给出色调，向量的长度是饱和度，在该平面中，所有彩色的强度由平面在垂直强度轴的位置给出。

从 RGB 到 HSI 的彩色转换

给定一幅 RGB 彩色格式的图像，每一个 RGB 像素和 H 分量可用下面的公式得到：

$$H = \begin{cases} \theta & B \leqslant G \\ 360 - \theta & B > G \end{cases} \tag{6.2.2}$$

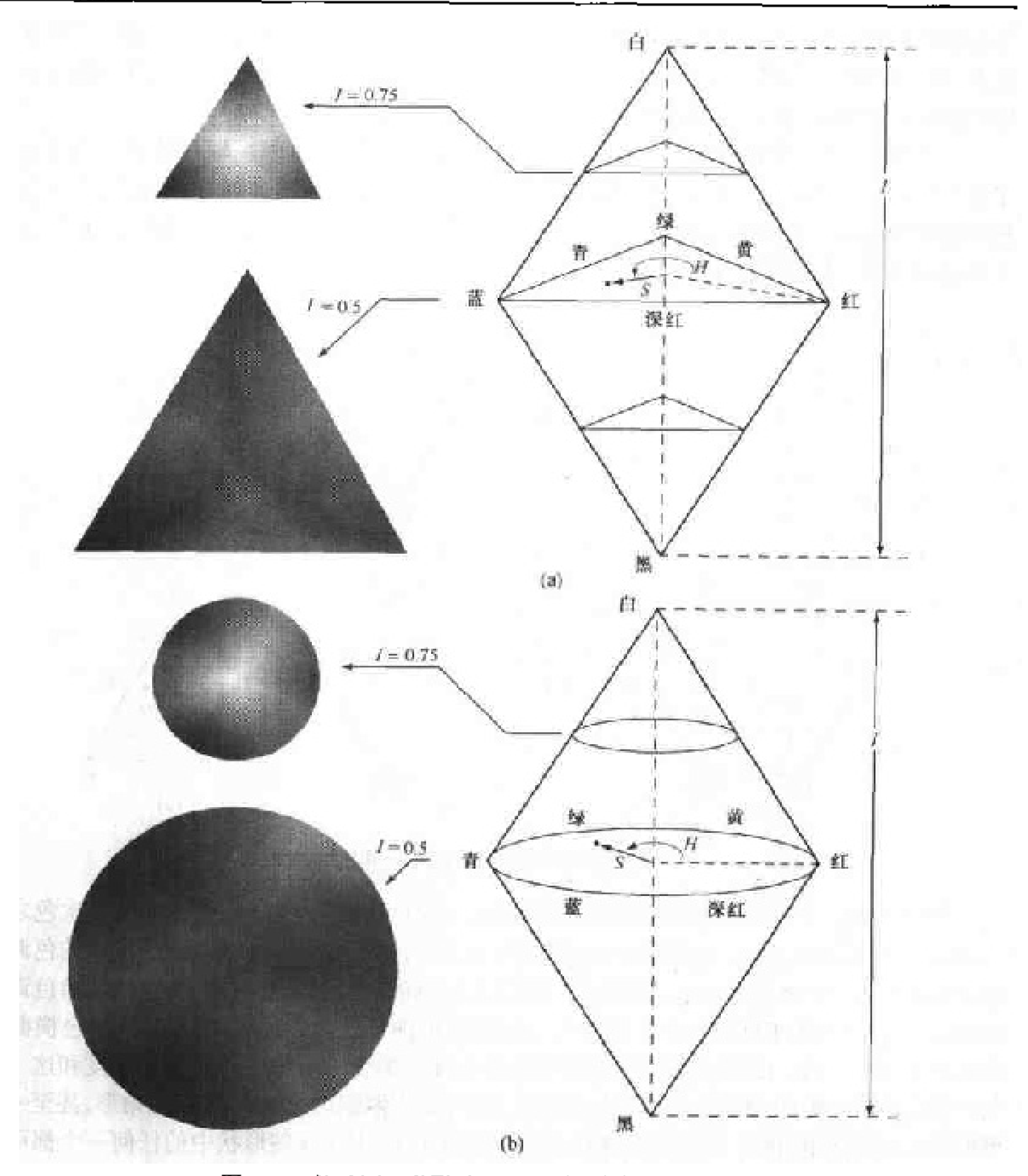


图 6.14 基于(a)三角形平面和(b)圆形彩色平面的 HSI 彩色模型。三角形和圆形平面垂直于强度轴

此处

$$\theta = \arccos\left\{\frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-G)(G-B)]^{1/2}}\right\}$$

色饱和度分量由下式给出：

$$S = 1 - \frac{3}{(R+G+B)}[\min(R,G,B)] \tag{6.2.3}$$

最后,强度分量由下式给出:

$$I = \frac{1}{3}(R + G + B) \tag{6.2.4}$$

如图 6.13 指出的那样,假定 RGB 值归一化为[0,1]范围内,角度 θ 根据HSI空间的红轴来度量。色调可以用式(6.2.2)得到的值除以 360°归一化为[0,1]范围内。如果给出的RGB值在[0,1]区间内,则其他两个HSI分量已经在[0,1]范围内了。

式(6.2.2)到式(6.2.4)的结果可以从图 6.12 和图 6.13 所示的几何关系推得。这一推导很繁琐,而且对现有的讨论没有太大意义。感兴趣的读者可参考有关书籍或访问证明这些公式及有关HSI到RGB转换的Web站点。

从 HSI 到 RGB 的转换

在[0,1]内给出 HSI 值,现在要在相同的值域找到RGB值,可利用 H 值公式。在原始色分割中有 3 个相隔 120°的扇形(见图 6.13)。从 H 乘以 360°开始,这时色调值返回原来的[0°, 360°]的范围。

***RG* 扇形($0° \leq H < 120°$)**:当 H 位于这一扇形区时,RGB 分量由下式给出:

$$B = I(1 - S) \tag{6.2.5}$$

$$R = I\left[1 + \frac{S\cos H}{\cos(60° - H)}\right] \tag{6.2.6}$$

$$G = 1 - (R + B) \tag{6.2.7}$$

***GB* 扇区($120° \leq H < 240°$)**:如果给定的 H 值在这一扇区,首先从 H 中减去 120°,即:

$$H = H - 120° \tag{6.2.8}$$

然后 RGB 分量为:

$$R = I(1 - S) \tag{6.2.9}$$

$$G = I\left[1 + \frac{S\cos H}{\cos(60° - H)}\right] \tag{6.2.10}$$

$$B = 1 - (R + G) \tag{6.2.11}$$

***BR* 扇区($240° \leq H < 360°$)**:最后,如果 H 在这一扇区,从 H 中减去 240°:

$$H = H - 240° \tag{6.2.12}$$

然后 RGB 分量为:

$$G = I(1 - S) \tag{6.2.13}$$

$$B = I\left[1 + \frac{S\cos H}{\cos(60° - H)}\right] \tag{6.2.14}$$

$$R = 1 - (G + B) \tag{6.2.15}$$

在图像处理中这些公式的用法将在下面各节加以讨论。

例 6.2 对应于 RGB 立方体的图像的 HSI 值

图 6.15 显示了图 6.8 中所示的 RGB 图像的色调、饱和度和强度。图6.15(a)是色调图像，其最大的特点是，在立方体前平面(红的)沿 45°线的值是不连续的。为了理解不连续的原因，参见图 6.8，从立方体的红色顶点到白色顶点画一条线，在这条线的中间选一点，从该点开始，向右环绕立方体一周直到回到开始点画一条轨迹。在该路径中遇到的主要颜色是黄、绿、青、蓝、深红、黑直至红。根据图 6.13，沿着这条路径的色调值应该从 0°到 360°增加(即从色调的最低到最高的可能值)。这正好如图6.15(a)所示的那样，因为在灰度级中最低值代表黑，最高值代表白。事实上，色调图像原本被归一化为[0,1]范围内，并标为 8 比特，即在显示时被转换为[0,255]范围的值。

在图6.15(b)中，饱和度图像显示出从暗值渐进地向RGB立方体的白顶点过渡，表明彩色饱和度越来越小，一直逼近白色。最后，示于图6.15(c)的强度图像中的每一个像素值等于图 6.8 中相应像素RGB的平均值。

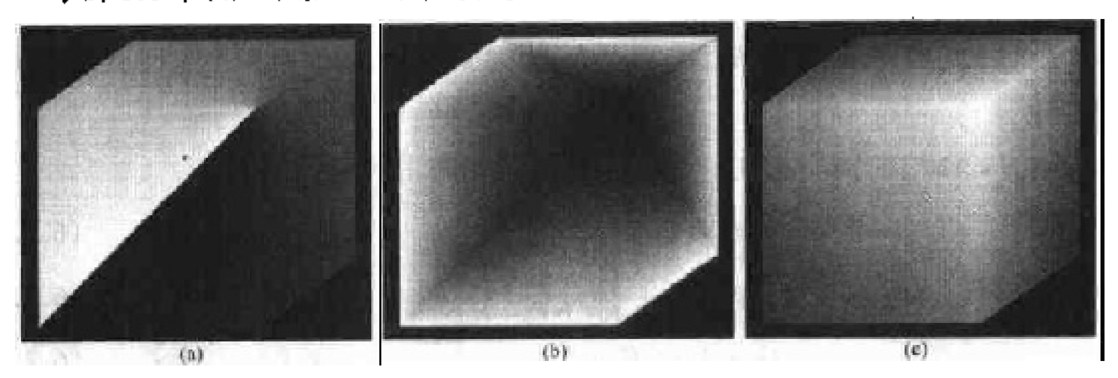


图 6.15 图 6.8 中的图像分量。(a)色调，(b)饱和度，(c)强度

HSI 分量图像处理

在下面的讨论中，考察 HSI 分量图像处理的几个简单技术，以便帮助熟悉这些分量，并可以加深对HSI彩色模型的理解。图 6.16(a)显示了一幅由 RGB 基色和二次色合成的图像。图 6.16(b)到(d)显示了这幅图像的 H, S 和 I 分量。这些图像是用式(6.2.2)到式(6.2.4)产生的。回顾一下本章早些时候的讨论：图6.16(b)中灰度值对应于角度；例如，因为红色对应 0°，则图6.16(a)中的红区域映射到色调图像中的黑区域。类似地，图6.16(c)中的灰度级对应饱和度(显示时划分等级为[0,255])，图6.16(d)中的灰度级是平均强度。

为了分别改变 RGB 图像任何区域的彩色，可以改变图6.16(b)色调图像中相应区域的值。然后与没有变化的 S 和 I 图像一起用式(6.2.5)到式(6.2.15)描述的步骤，把新的 H 图像变换回到RGB。为了在任何区域改变彩色的饱和度(纯度)，除了在HSI空间改变饱和度外可遵循相同的步骤。类似的讨论可用于改变任何区域的平均强度。当然，这些改变可同时进行。例如，图 6.17(a)的图像是把对应于图6.16(b)的蓝、绿区域的像素变为 0 得到的。图6.17(b)的图像是图 6.16(c) 分量图像 S 中青色区域的饱和度减少一半得到的图像。图 6.17(c) 是把图 6.16(d)的强度图像中心的白区域强度减少一半的图像。把改进的HSI图像变换回RGB的结果显示在图6.17(d)中。正如预期的那样，在这些图中可以看到，所有圆形的外部区域是红的，青区域的强度减小，中心区变灰而不是白色。显然这些结果很简单，却显示出HSI模型独立控制色调、饱和度和强度的能力，当描述彩色时，会十分熟悉这些量。

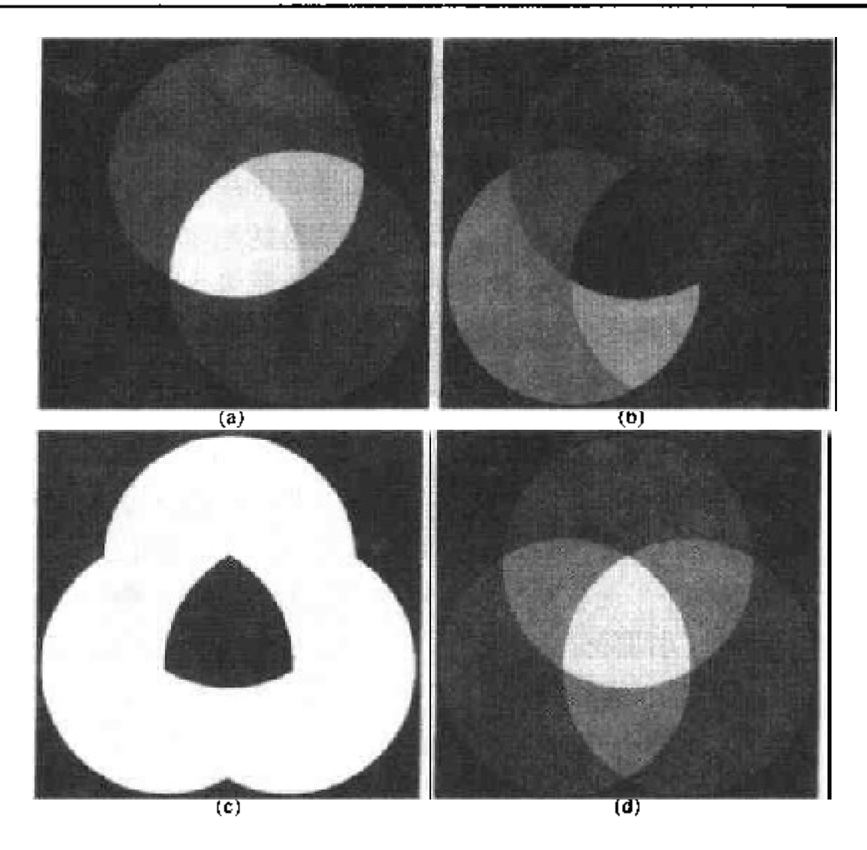


图 6.16 (a)RGB 图像和与之相应的HSI图像分量,(b)色调,(c)饱和度,(d)强度

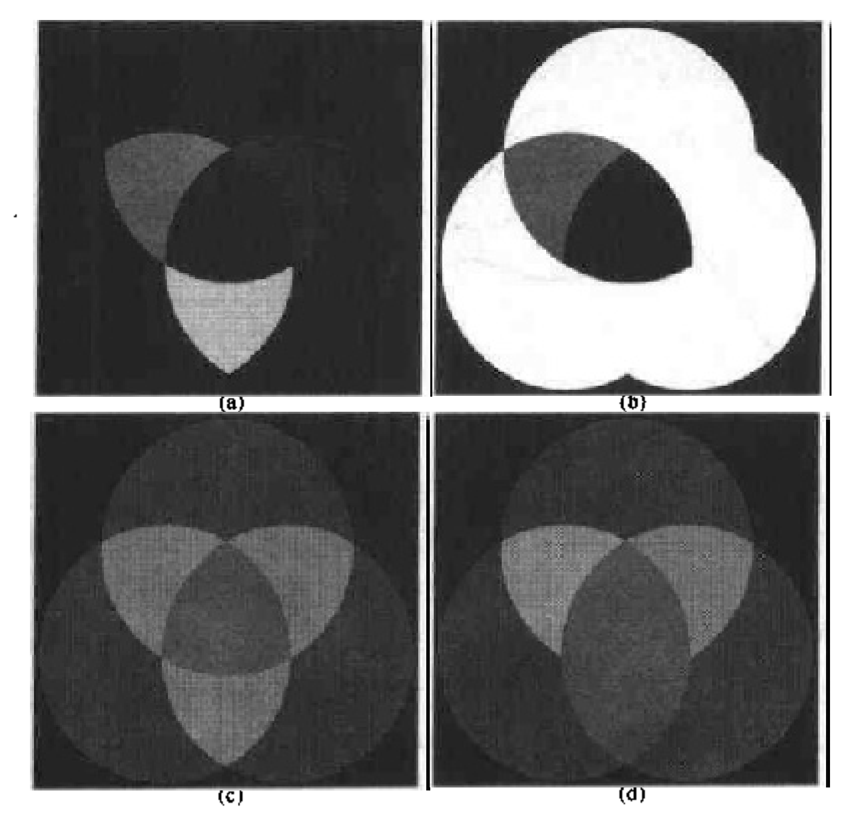


图 6.17 (a)到(c)改变的 HSI 分量图像,(d)最终的 RGB 图像

6.3 伪彩色处理

伪彩色图像处理(也称假彩色)是根据特定的准则对灰度值赋以彩色的处理。伪彩色或假彩色一词用于区分真彩色图像处理和为单色图像赋以彩色的处理。这种处理是6.4节将讨论的题目。伪彩色的主要应用是为了人目视观察和解释一幅图像或序列图像中的灰度目标。正像在本章开始指明的那样,利用彩色的主要动力之一是,人类可以辨别上千种颜色和强度,而相形之下只能辨别二十几种灰度。

6.3.1 强度分层

强度分层技术(有时也叫密度分层)和彩色编码是伪彩色图像处理最简单的例子之一。如果一幅图像被描述为三维函数(作为空间坐标的强度),分层方法可看成是放置一些平行于图像坐标面的平面,然后每一个平面在相交的区域中切割图像函数。图6.18显示了利用平面把图像函数 $f(x,y)=l_i$ 切割为两部分。

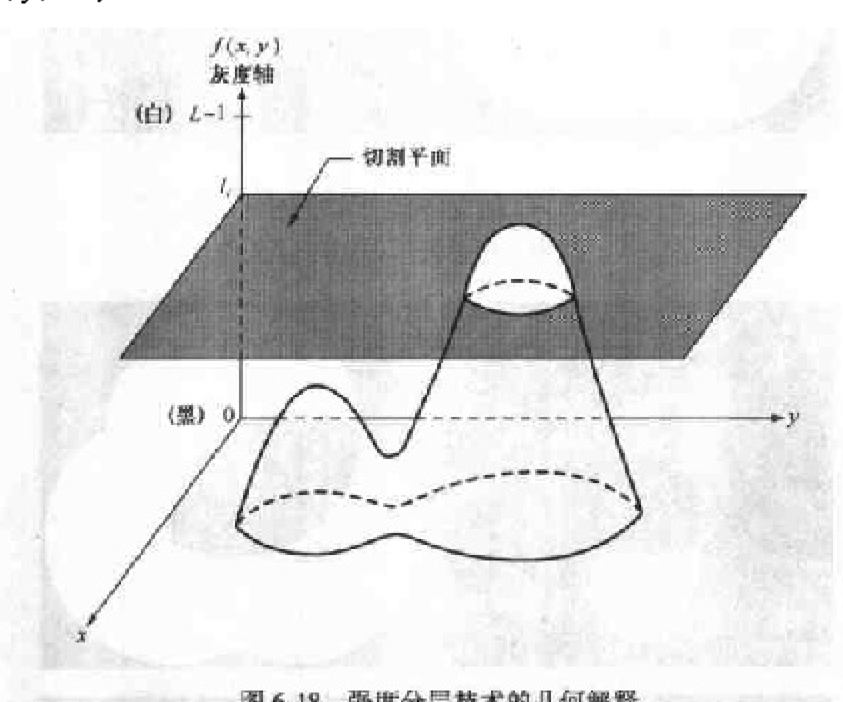


图6.18 强度分层技术的几何解释

如果对图6.18所示的平面每一面赋以不同的彩色,平面上任何灰度级上的像素将编码成一种彩色。任何平面之下的像素将编码为另一种彩色。位于平面上的灰度级本身被任意赋以两种彩色之一。其结果是一幅两色图像,相关的彩色状态可由沿灰度轴上下移动切割平面来控制。

通常,该技术可总结如下:令$[0,L-1]$表示灰度级(见2.3.4节),使 l_0 代表黑色$[f(x,y)=0]$,l_{L-1}代表白色$[f(x,y)=L-1]$。假定垂直于强度轴的 P 个平面定义为量级 $l_1,l_2,\cdots,l_P$。并假定,$0<P<L-1$,P 个平面将灰度级分为$P+1$ 个间隔,$V_1,V_2,\cdots,V_{P+1}$。灰度级到彩色的赋值根据如下关系进行:

$$f(x,y)=c_k \qquad f(x,y)\in V_k \tag{6.3.1}$$

这里 c_k 是与强度间隔 V_k 第 k 级强度有关的颜色，V_k 是由在 $l=k-1$ 和 $l=k$ 分割平面定义的。

平面的概念对于强度分层技术的几何解释很有用。图 6.19 显示了另一种解释，它定义了如图 6.18 所示的相同的映射。根据图 6.19 所示的映射函数，任何输入的灰度级根据其位于 l_i 值的上面还是下面而赋于两种颜色之一。当用更多的灰度级时，映射函数采取阶梯形式。

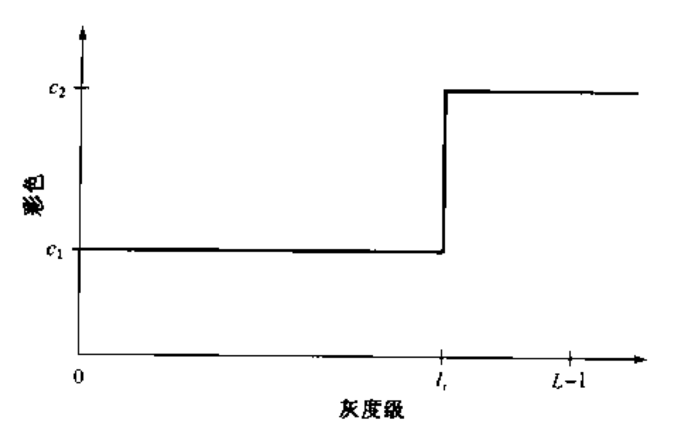


图 6.19　强度分层技术的另一种解释

例 6.3　强度分层

一个简单而实用的强度分层法示于图6.20。图6.20(a)是Picker甲状腺模型(放射实验模型)的单色图像，图6.20(b)是强度分层结果。图像分为 8 个彩色区域。在单色图像中出现的恒定强度区域也是完全可变的，如在分区图像中变化的颜色所示。例如左瓣在单色图像中是暗灰度，以强度取出病变相对很困难，彩色图像清楚地显示恒定强度的不同区域，每一区域采用一种颜色。

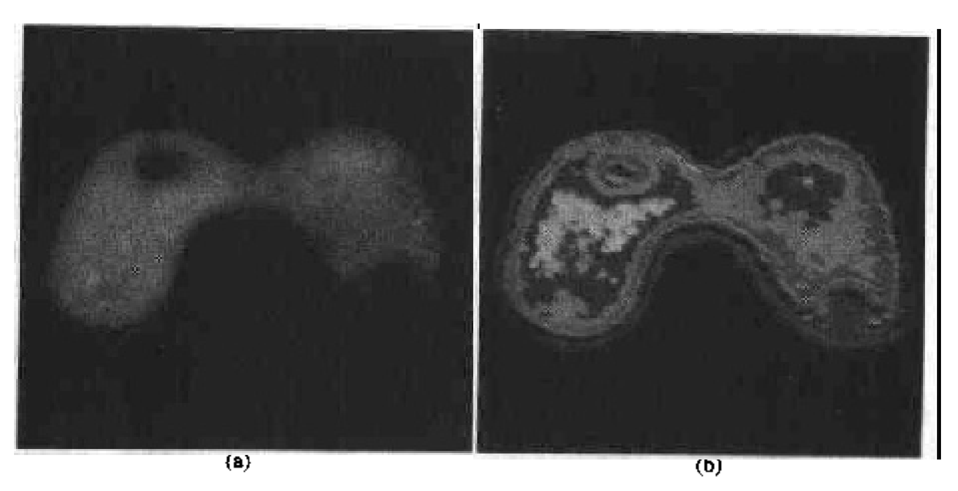
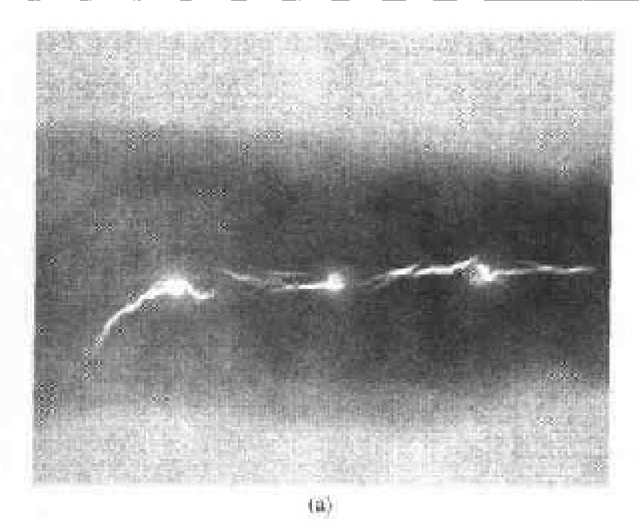
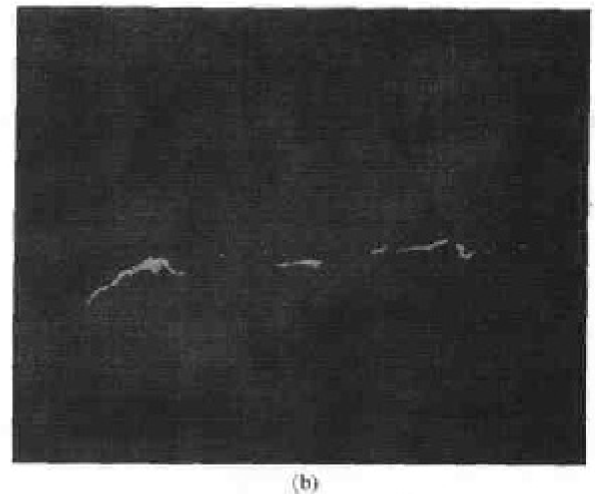


图 6.20　(a)Picker 甲状腺模型的单色图像，(b)强度分为 8 个彩色的结果（Oak Ridge 国家实验室仪器与控制分部提供）

在以上简单例子中，灰度被分成间隔并对每一区域赋以不同的颜色，在图像中并没有涉及灰度级的意义。在那种情况下，感兴趣的是简单地观察指定图像的灰度级。当基于图像的物理特性细分灰度级时，强度分层有很大意义并承担重要角色。例如图6.21(a)所示的含有裂纹和孔隙(通过图像中部水平连续的亮白条痕)的焊接物 X 光图像(水平暗区)。



(a)



(b)

图 6.21 (a)焊接的单元色 X 光图像,(b)彩色编码结果(原图像由 X-TEK 系统公司提供)

我们知道当焊接中有裂痕和气孔时,X 射线的全部强度都通过物体使位于物体另一边的成像传感器饱和(见 2.3 节)。这样,来自该系统的8 比特图像中,图像灰度为 255,意味着焊点有问题。如果人是最终的分析判断者,并采用人工检测焊接质量(直至今天还是常用的工序),简单的彩色编码给 255 灰度级赋以一种彩色,其他的所有灰度赋以另一颜色就会极大地简化检测者的工作。图6.21(b)显示了这一结果。如果以图6.21(b)的形式代替图6.21(a)显示图像,"人的误识率将降低"这一结论毋庸置疑。换句话说,如果灰度级的确切值是已知的,灰度分层在可视化方面是简单而有力的手段。特别是涉及大量图像时更是如此。下面先以一个有意义和更复杂的例子来加以说明。

例 6.4 用颜色突出降雨水平

测量降雨水平,特别在地球热带区域与环境有关的多种应用中是很重要的。因地面传感器难于准确测量而且代价昂贵,得到总雨量图更是困难(因为重要的降雨区发生在海上)。获得降雨图的一个方法是使用卫星。TRMM(热带降雨测量任务)卫星使用为检测降雨特殊设计的 3 个传感器:一个降雨雷达、一个微波成像器和一个可见光及红外扫描仪(见1.3 节和 2.3 节谈到的图像传感方式)。

来自各种雨量传感器的结果经处理,得出由传感器监视的区域在给定时间区段平均降雨的估计,从这些估计,得到灰度级图像并不困难。图像的强度值直接与降雨相对应,每一个像素代表实在的陆地面积,其大小取决于传感器的分辨率。这样一幅强度图像示于图 6.22(a),这里由卫星监视的区域在图像的中间三分之一处是稍亮一些的水平带(这就是降雨区)。在这个特殊例子中,降雨值是在三年时间内一个月的平均值[以英寸(in)为单位,1 in = 0.0254 m]。

即使可能,对降雨模式,目测这些图像也是十分困难的。然而,假设使用图6.22(b)所示的彩色对 0 到 255 灰度级编码,趋于蓝色值意味着低降雨值(量),相反则为红色。标度终止于纯红色处,表示降雨量大于 20 in。

图6.22(c)显示了用刚才讨论的彩色图对灰度图像进行彩色编码的结果。如图6.22(c)和图6.22(d)的放大图所示,这一结果容易理解得多。另外,除了覆盖全球范围,这种类型的数据允许气象工作者用比以往更精确的方法标定基于地面的雨量监测系统。

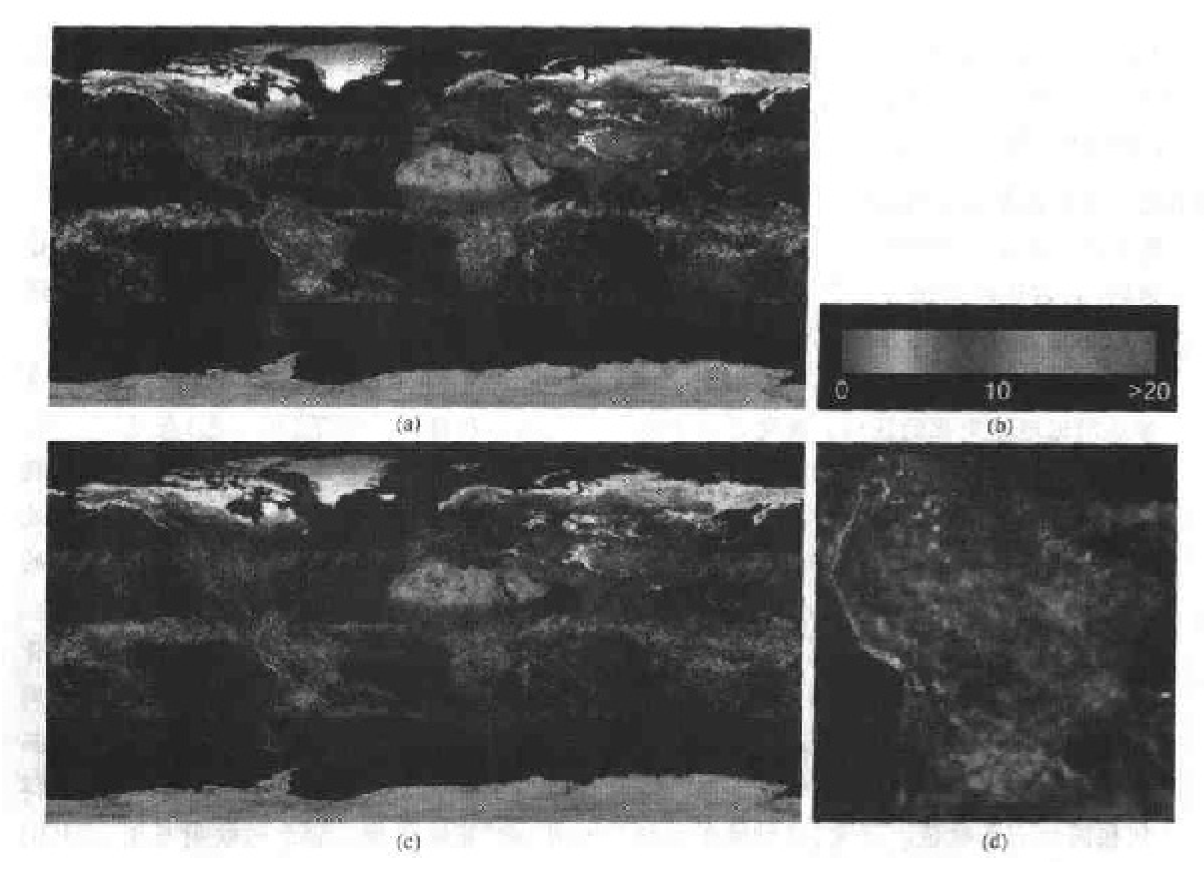


图 6.22　(a)强度与月平均降雨相对应的灰度图像，(b)对强度值赋予的彩色，(c)彩色编码图像，(d)南美区域的放大图(NASA提供)

6.3.2　灰度级到彩色转换

与前一节讨论的简单分层技术比较，其他类型的变换更通用也更能拓宽伪彩色增强结果的范围。一种特别有吸引力的方法示于图 6.23。这一方法的基本概念是对任何输入像素的灰度级执行 3 个独立的变换。然后，3 个变换结果分别送入彩色电视监视器的红、绿、蓝通道。这种方法产生一幅合成图像，其彩色内容受变换函数特性调制，注意，这些方法是一幅图像灰度值的变换而不是位置函数。

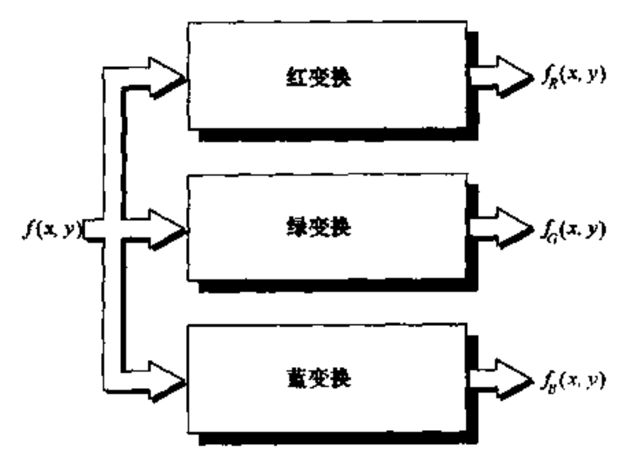


图 6.23　伪彩色图像处理的能力框图。f_R、f_G、f_B 是送入 RGB彩色监视器的相对应的红、绿、蓝输入信号

前面各节讨论的方法仅仅是描述该技术的特殊情况,用以产生彩色的是灰度级的分段线性函数(图6.19)。另一方面,这一节讨论的方法是基于平滑的非线性函数,正如所希望的那样,这种方法有相当大的技术上的灵活性。

例6.5 突出装在行李内的爆炸物的伪彩色应用

图6.24(a)显示了由机场X光扫描系统得到的行李的两幅单色图像。左边的图像含有普通物品,右边的图像含有相同的物品以及一块模拟的塑料爆炸物。这个例子的目的是说明应用灰度级到彩色的变换可以得到各种增强程度。

图6.25显示了所用的变换函数。这些正弦函数包含在峰值附近有相对恒定值的区域和在谷底附近迅速变化的区域。改变每个正弦函数的相位和频率强调了(用彩色)灰度级范围。例如,如果所有的3个变换都有相同的相位和频率,则输出图像将是单色的。3个变换函数间相位上的一个微小变化可产生对应正弦峰值的灰度像素的一个微小变化,特别是正弦波有较宽曲线时更是如此(低频情况)。由于正弦波间的相位移导致3个正弦波幅度间显著的不同,位于正弦波徒峭部分的那些灰度级的像素被赋予强得多的彩色内容。

示于图6.24(b)的图像是用图6.25(a)中的变换函数得到的,它显示了与爆炸物、衣服和背景分别相对应的灰度范围。注意,爆炸物和背景有很不同的灰度级,但是由于正弦波的周期性,它们都用近似相同的颜色编码。示于图6.24(c)的图像由图6.25(b)所示的变换函数得到。在这种情况下,爆炸物和服装袋的强度范围由相似的变换映射,这样,基本上得到相同的彩色赋值。注意,这种映射允许观测者"看"穿爆炸物。背景的映射与图6.24(b)所用的相同,产生几乎相同的彩色赋值。

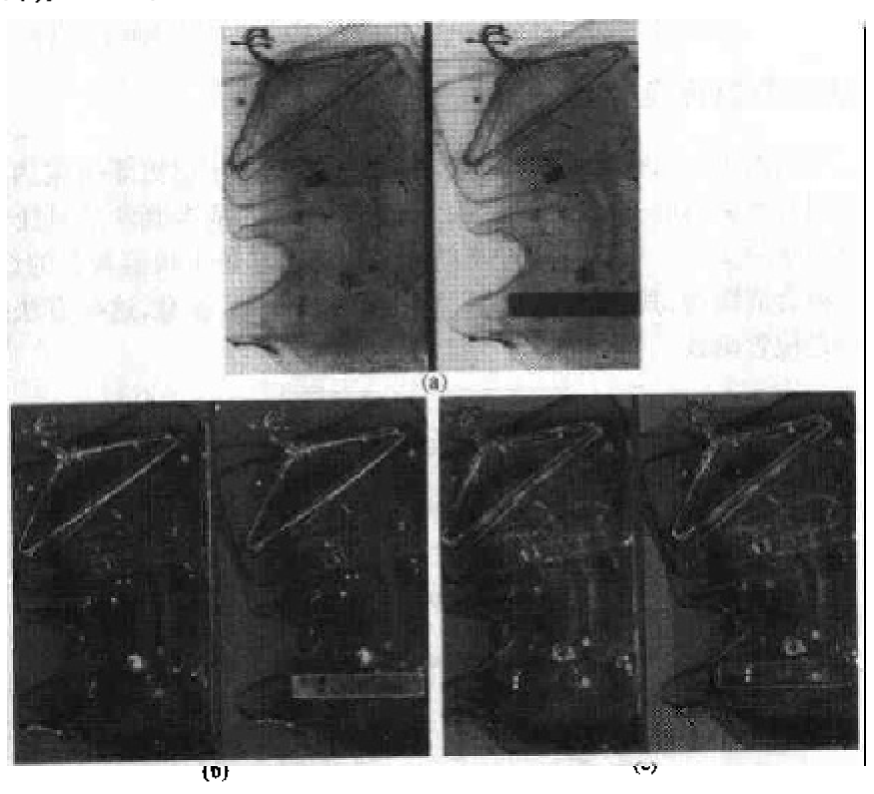


图6.24 使用图6.25灰度级到彩色变换的伪彩色增强(原图像由Westing house公司的Mike Hurwitz博士提供)

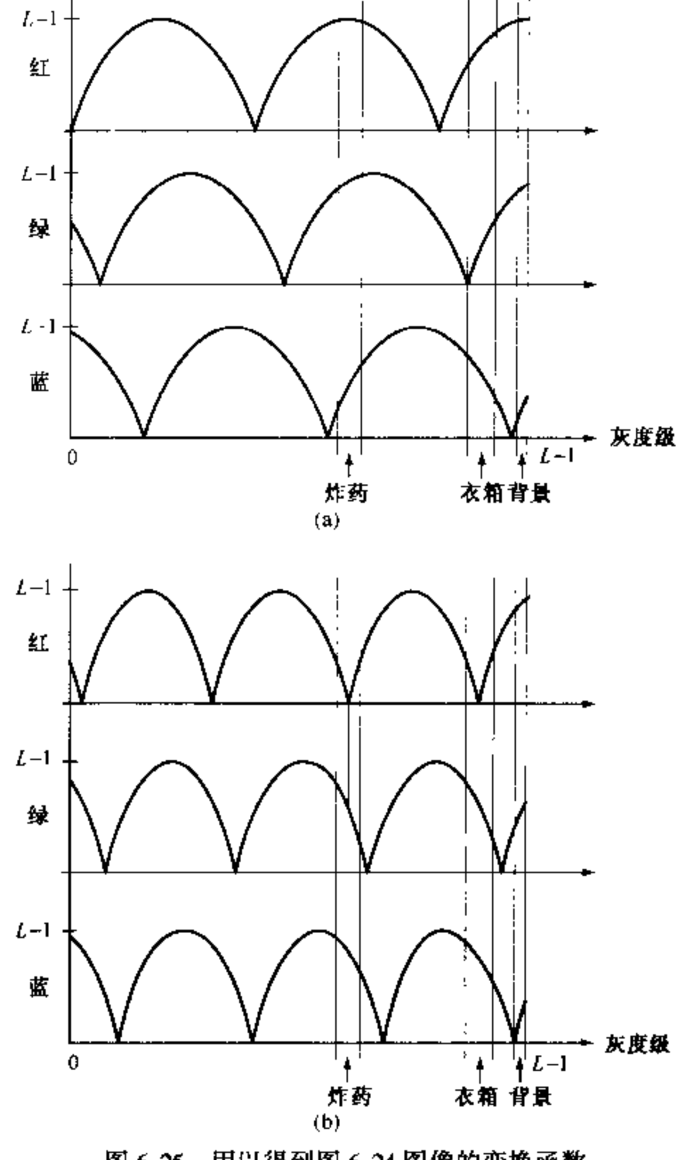


图 6.25　用以得到图 6.24 图像的变换函数

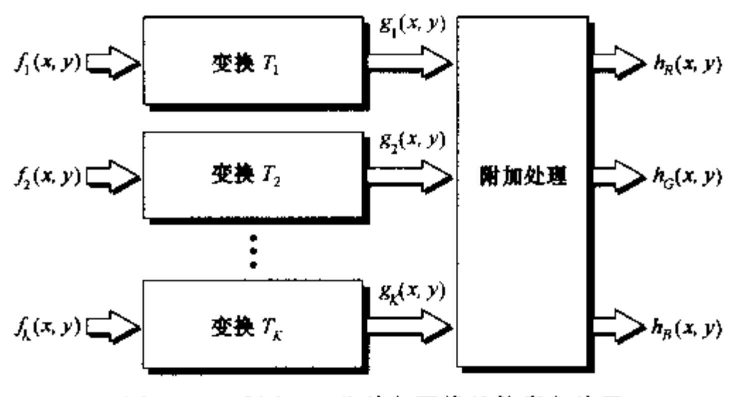


图 6.26　适用于一些单色图像的伪彩色编码

图 6.23 所示的过程以一幅单色图像为基础。通常如图 6.26 所示，将多幅单色图像组合为一幅彩色组合图像是很重要的。在多光谱图像处理中常使用这一过程。这里，不同的传感器在不

同的谱段产生独立的单色图像。图 6.26 所示的附加处理类型可以是彩色平衡(见 6.5.4 节)混合图像，选择三幅用于显示的图像，该图像基于产生图像的传感器的响应特性的知识。

例 6.6　多光谱图像彩色编码

图 6.27(a)到(d)显示了 4 幅华盛顿的光谱卫星图像，包括波托马克河部分。前 3 幅是可见光的红、绿、蓝图像，第 4 幅是近红外图像(见表 1.1 和图 1.10)。图6.27(e)是把前 3 幅图像合成为RGB图像得到的全彩色图像。密度大的全彩色图像区域很难判读，但该图像显著的特点是波托马克河的各部分颜色不同。图6.27(f)较有趣，这幅图像是由近红外图像代替图6.27(e)的红分量形成的。从表1.1可知这一波段对场景的生物分量有较强反应。图 6.27(f)显示生物(红色)和场景中人造目标的特性间有十分明显的差别，由混凝土和柏油组成的部分在图像中呈现浅蓝色。

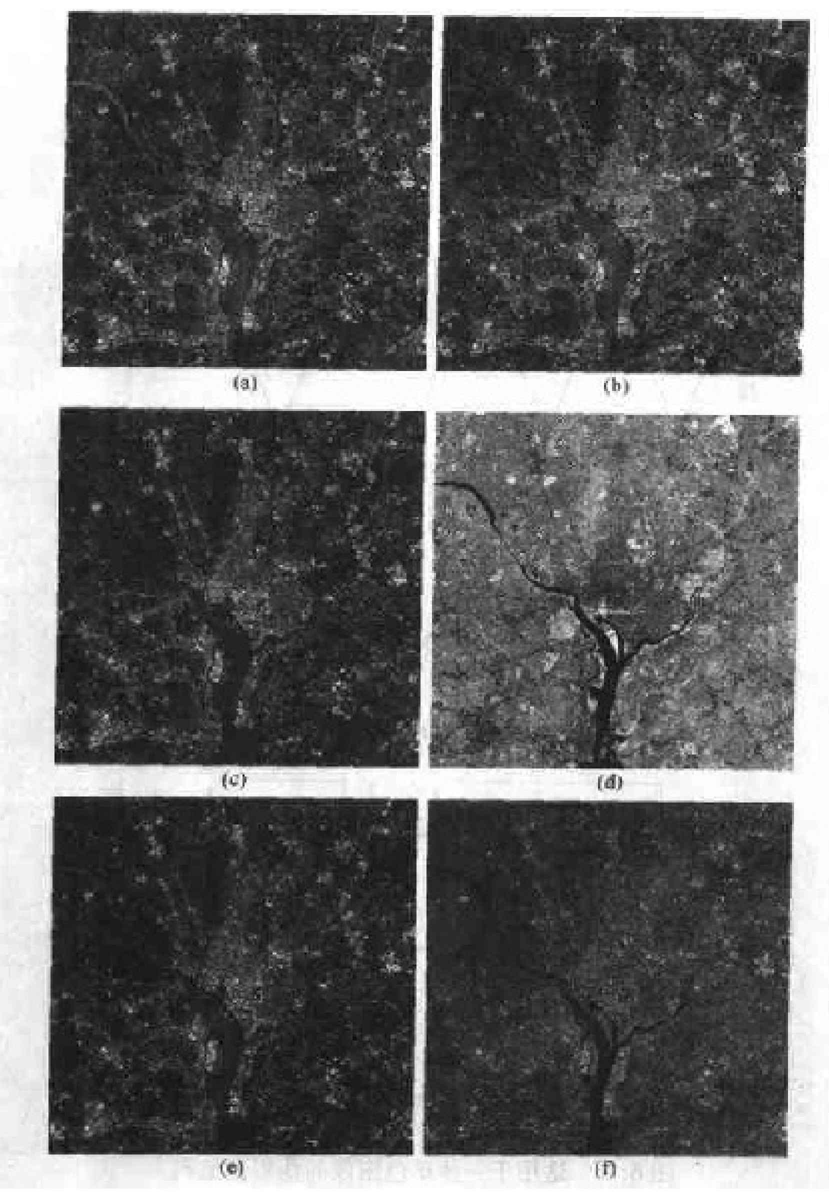
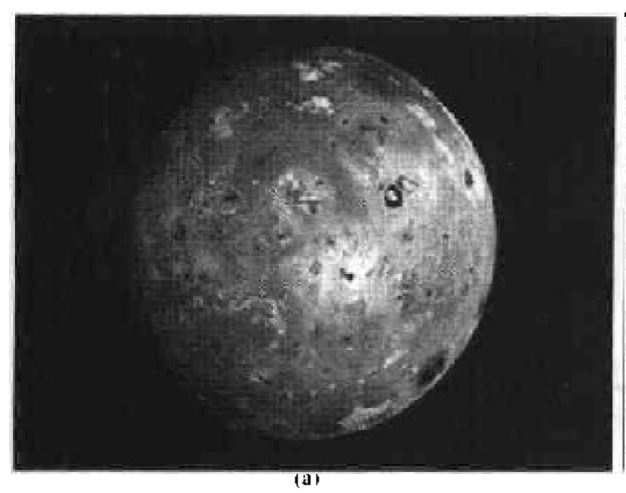


图 6.27　(a)～(d)是图 1.10 中的 1 到 4 波段的图像(见表 1.1)，(e)以 RGB 图像红、绿、蓝分量处理图(a)，(b)，(c)得到的彩色合成图像，(f)用相同的方法，但在红通道使用(d)中的近红外图像得到的图像(原多光谱图像由 NASA 提供)

刚才说明的处理类型在复杂图像中可视化感兴趣事物时功能很强，特别是那些超出我们正常感知能力之外的事物。图 6.28 对这一点是很好的说明。这是木星图像，这些图像是把来自Galileo宇宙飞船的几幅传感器图像组合并以伪彩色显示，其中一些是人眼不可见的谱范围。然而，在类似于改变传感器响应的物理和化学过程中，有可能把感知的图像组合为有意义的伪彩色图像。混合图像数据的一个方法在于如何显示化学成分表面的差别，或者表面反射太阳光方式的改变。例如，图6.28(b)中的伪彩色图像，亮红色表示从活火山所喷出的物质，而周围黄色物质是陈旧的硫沉积物。这些图像传递这些特性要比单独分析分量图像快得多。


(a)

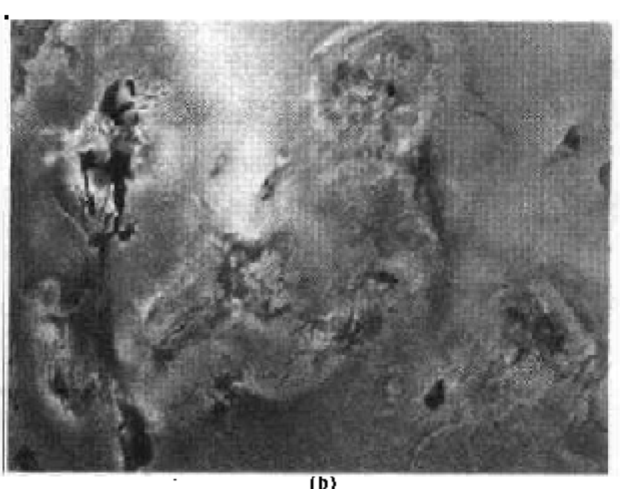

(b)

图 6.28　(a)木星的伪彩色图像，(b)靠近的图像(由 NASA 提供)

6.4　全彩色图像处理基础

这一节将讨论用于全彩色图像的处理技术。虽然该技术的研究远没有穷尽，但本节稍后阐述的技术是直观的，即面对各种各样的图像处理任务怎样处理全彩色图像。全彩色图像处理研究分为两大类。第一类是分别处理每一分量图像，然后，从分别处理过的分量图像形成合成彩色图像。第二类是直接对彩色像素处理。因为全彩色图像至少有 3 个分量，彩色像素实际上是一个向量。例如，在RGB系统中，每一个彩色点可以用从RGB坐标系统原点延伸的向量来解释(见图 6.7)。

令 **c** 代表 RGB 彩色空间中的任意向量：

$$\mathbf{c} = \begin{bmatrix} c_R \\ c_G \\ c_B \end{bmatrix} = \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{6.4.1}$$

这个公式指出：**c** 的分量是一幅彩色图像在一点上的 RGB 分量。可以这样认为，彩色分量是坐标(x, y)的函数，表示为：

$$\mathbf{c}(x, y) = \begin{bmatrix} c_R(x, y) \\ c_G(x, y) \\ c_B(x, y) \end{bmatrix} = \begin{bmatrix} R(x, y) \\ G(x, y) \\ B(x, y) \end{bmatrix} \tag{6.4.2}$$

对大小为 $M\times N$ 的图像,有 MN 个这样的向量,$\mathbf{c}(x,y)$,$x=0,1,2,\cdots,M-1$;$y=0,1,2,\cdots,N-1$。

为清楚起见,考虑式(6.4.2)描述的向量是很重要的,这些分量是 x 和 y 的空间变量。这是一个常常产生混乱的根源,可以把焦点放在空间处理上来避免混淆。这就是我们感兴趣的用 x 和 y 公式化的图像处理技术。现在,像素是彩色像素这一事实引入一个因子,在其最简单的公式中允许用标准的灰度图像处理方法去分别处理彩色图像的每一分量。然而,单独的彩色分量的处理结果并不总等同于在彩色向量空间的直接处理,在这种情况,就必须采取新的公式化方法。为了使每一彩色分量处理和基于向量的处理等同,必须满足两个条件:第一,处理必须对向量和标量都可用,第二,对向量每一分量的操作对于其他分量必须是独立的。正如插图所示,图6.29显示了邻域灰度空间处理和全彩色处理。假定该处理是邻域平均,在图6.29(a)中,平均是将邻域内的所有像素灰度级相加然后用邻域内像素总数去除。在图6.29(b)中,平均是把邻域内所有向量相加并用邻域内向量总数去除每一个分量。但平均向量的每一个分量是对应其分量的图像像素和。这与在每个彩色分量基础上做平均,然后形成向量得到的结果是相同的。在下面一节中将更详细地说明这一处理,还要示出一种方法,在这种方法中,两种处理的结果是不同的。

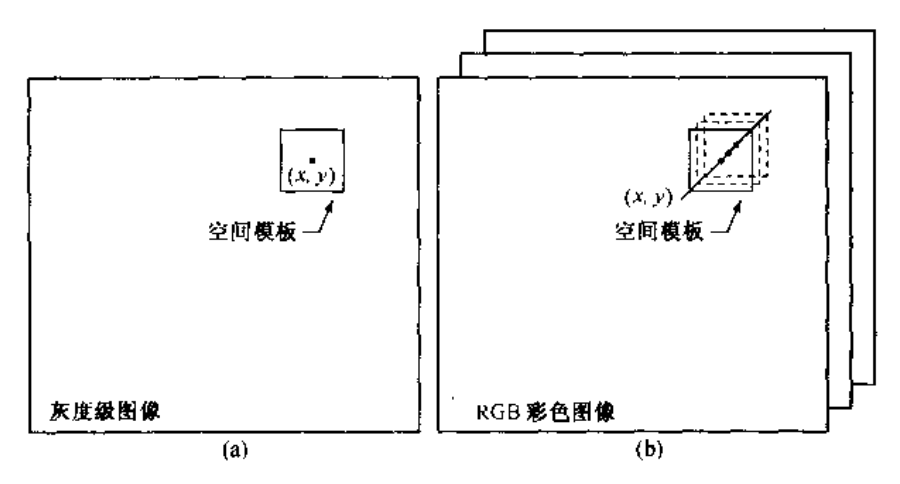


图 6.29 对灰度和 RGB 彩色图像的空间掩模

6.5 彩色变换

本节讨论的技术,总称为彩色变换,涉及在单一彩色模型的范围中处理彩色图像分量,而不是模型间的那些分量的转换(像 6.2.3 节的RGB到HSI,HSI到RGB的转换)。

6.5.1 公式

如第 3 章的灰度变换技术,我们用下式表达的彩色变换做模型:

$$g(x,y) = T[f(x,y)] \tag{6.5.1}$$

这里 $f(x,y)$是彩色输入图像,$g(x,y)$是变换或处理过的彩色输出图像,T 是在空间邻域(x,y)上对 f 的操作。这个公式和式(3.1.1)间的主要区别是在其描述上。这里,像素值是从彩色空

间选择的3元组或4元组,用来描述图像,如图6.29(b)所示。

类似于在 3.2 节用过的引进基本灰度变换的过程。在这一章仅限于彩色变换:

$$s_i = T_i(r_1, r_2, \cdots, r_n) \qquad i = 1,2,\cdots,n \tag{6.5.2}$$

这里,为标记简单,r_i 和 s_i 是$f(x,y)$和 $g(x,y)$在任何点处彩色分量的变量,$\{T_1, T_2, \cdots T_n\}$是一个对 r_i 操作产生 s_i 的变换或彩色映射函数集,注意到 n 个变换 T_i 合并执行式(6.5.1)中的单一变换函数 T。所选择的用于描述 f 和 g 像素的彩色空间决定 n 的值。如果选择RGB彩色空间,则 $n=3$, r_1, r_2 和 r_3 分别表示输入图像的红、绿、蓝分量。如果选择CMYK或HSI彩色空间,则 $n=4$ 或 $n=3$。

图 6.30(a)显示了一碗草莓和一个咖啡杯的高分辨率彩色图像。这是从大幅(4″×5″)彩色负片数字化的图像。图中的第二行包含原始的CMYK扫描分量图像。在这些图像的每一个CMYK彩色分量中,白用 1 表示,黑用 0 表示。这样,我们看到草莓是由大量的深红和黄色组成的,因为对应于这两种CMYK分量的图像最亮。黑色较少并通常限于咖啡和草莓碗中的阴影。当CMYK图像被转换为RGB时,如图中第三行所示,可以看到草莓包含大量的红色和很少的绿色与蓝色。图 6.30 的最后一行显示了用式(6.2.2)到式(6.2.4)计算出的图6.30(a)的HSI分量图像。如期望的那样,强度分量是全彩色原像的单色复现。另外,草莓在彩色方面相对较纯净。它们具有最高的饱和度或图像中色调被白光稀释得最少。最后注意到说明色度分量时的某些困难。问题包含这样一些事实:(1)在HSI模型中,0°和 360°相遇处有一个不连续点;(2)色调对于 0 饱和度没定义(对白、黑和纯灰)。模型的不连续点多出现在草莓周围,它们用接近白(1)和黑(0)的灰度值描述。其结果是不希望的高对比灰度级的混合去描述单一颜色——红色。

图 6.30 中的任何彩色空间分量都可与式(6.5.2)相配合。理论上,任何变换都可在任何模型中执行。然而,实际上,某些操作对特定的模型比较适用。对给定的变换,表达式之间转换的代价必须转换为与彩色空间有关的判决,然后在该空间执行该转换。例如,假设要改进图 6.30(a)图像的亮度,使用

$$g(x,y) = kf(x,y) \tag{6.5.3}$$

这里 $0<k<1$。在 HSI 彩色空间,可以做简单变换:

$$s_3 = kr_3 \tag{6.5.4}$$

这里 $s_1=r_1$, $s_2=r_2$,仅仅HSI亮度分量 r_3 被改变。

在 RGB 彩色空间,3 个分量都必须变换:

$$s_i = kr_i \qquad i = 1,2,3 \tag{6.5.5}$$

CMY 空间要求一个相似的线性变换集:

$$s_i = kr_i + (1-k) \qquad i = 1,2,3 \tag{6.5.6}$$

虽然 HSI 变换包含最少的操作,把 RGB 或 CMYK 图像转换为 HSI 空间所要求的计算大大抵消了简单变换的优点,即转换计算量比亮度变换本身计算量大。然而,不考虑所选择的彩色空间,其输出是相同的。图6.31(b)显示了采用式(6.5.4)到式(6.5.6),对图6.30(a)用 $k=0.7$ 进行变换的结果。映射函数示于图6.31(c)到(e)。

注意式(6.5.4)到式(6.5.6)所定义的每一变换仅依赖于其彩色空间的一个分量,这一点很重要。例如,红的输出分量 S,在式(6.5.5)中独立于绿和蓝输入,它只依赖于红输入。正如在开始讨论提到的那样,这类变换是最简单的和最常用的彩色处理工具,并可以以每个彩色分

量为基础进行。在这一节余下的部分,我们研究几种这样的变换并讨论一种情况,在这种情况下,分量变换函数依赖所有的输入图像分量,因此,不能以单独彩色分量为基础做变换。

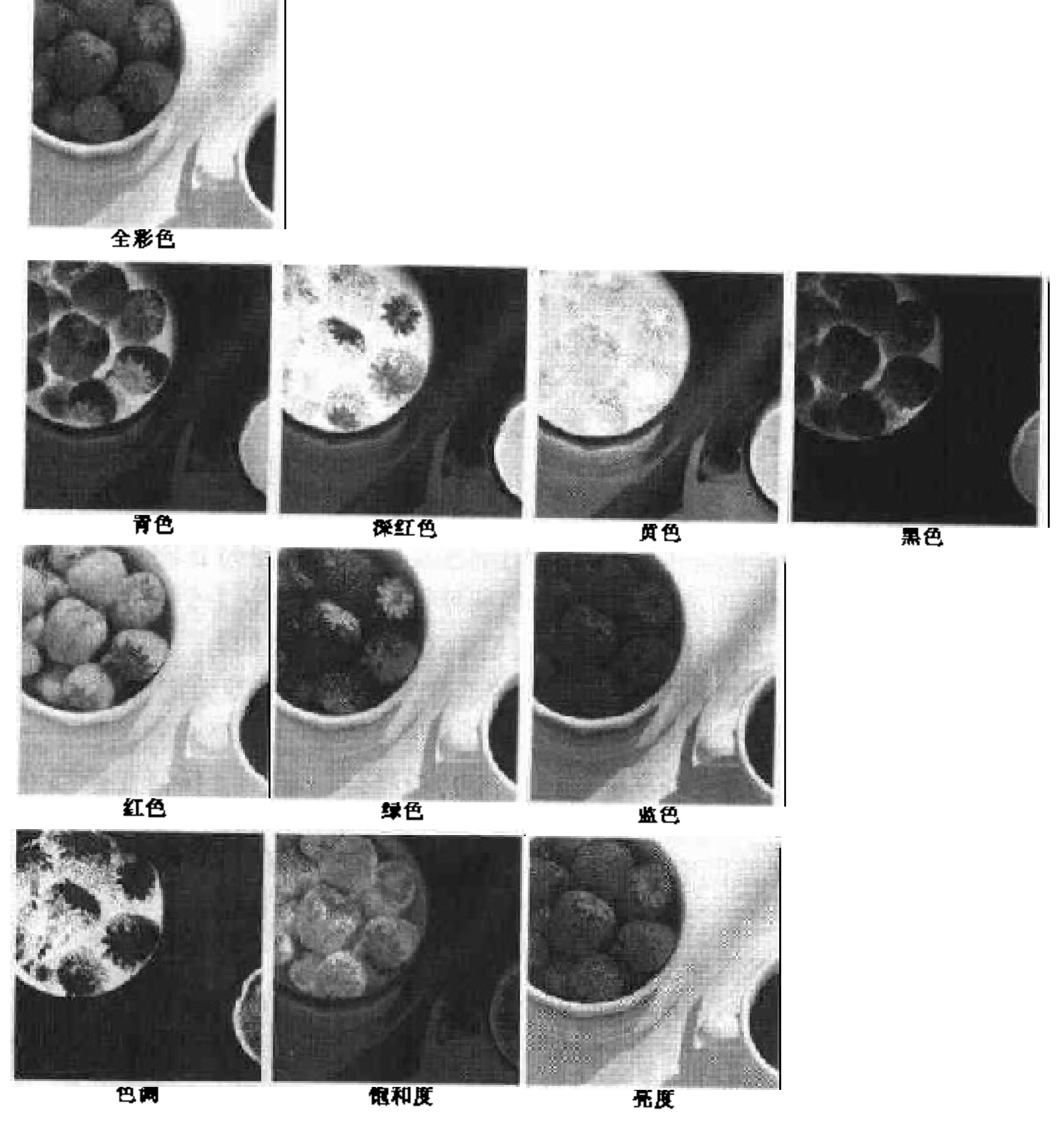


图 6.30 全彩色图像和它的不同彩色空间分量
(原图像由 MedData Interactive 提供)

6.5.2 补色

在图 6.32 的彩色环[①] 上,与一种色调直接相对立的另一种色调称为补色。对补色的兴趣源于一个事实,即补色类似于 3.2.1 节的灰度负值。正如在灰度情况下那样,补色对于增强嵌在彩色图像暗区的细节,特别是在大小上占支配地位的细节很有用。

① 由牛顿首创的彩色环,在 17 世纪他连接彩色谱的末端形成第一个彩色环。

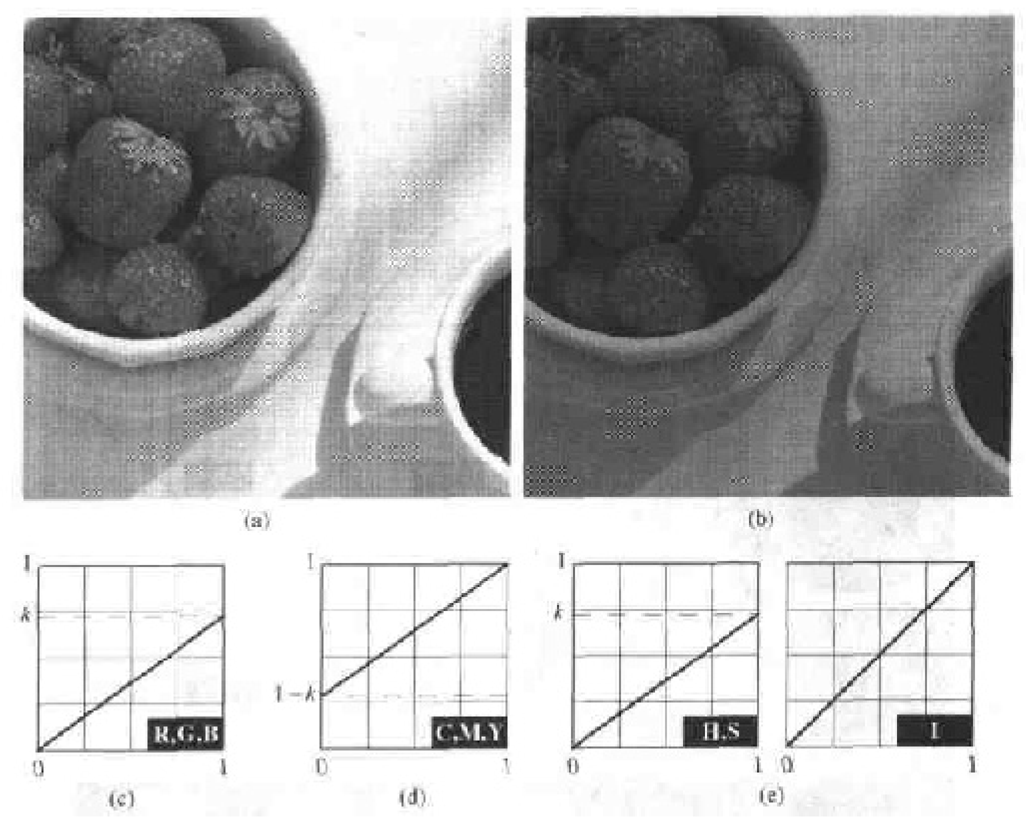


图 6.31　用彩色变换调整图像亮度。(a)原像,(b)减少亮度 30% 的结果($k=0.7$),(c)~(e)所要求的RGB,CMY和HSI 变换函数(图像由 MedData Interactive 提供)

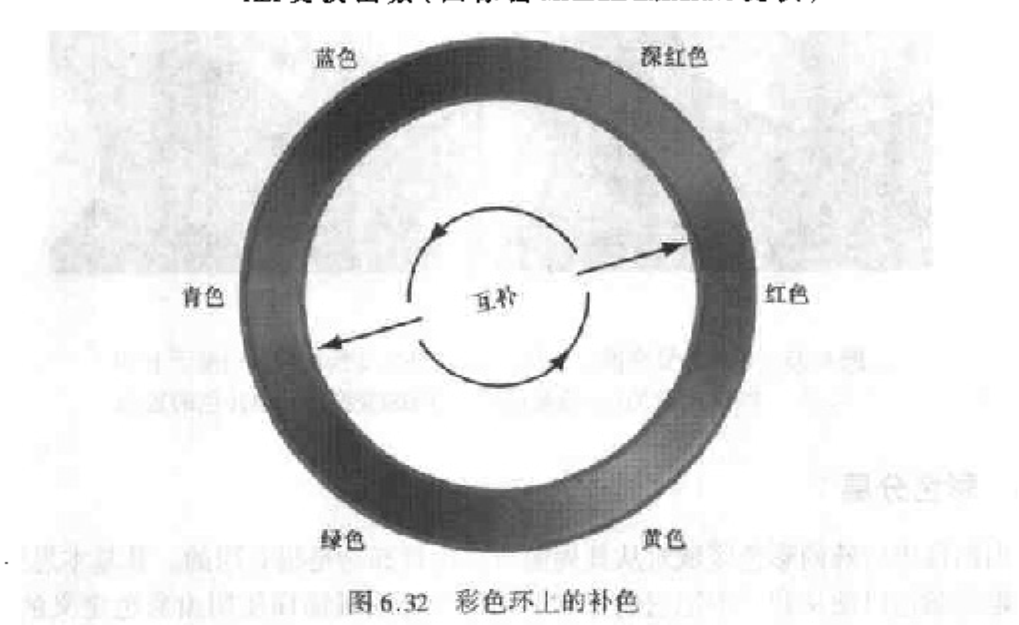


图 6.32　彩色环上的补色

例 6.7　计算彩色图像的补色

图 6.33(a)和(c)显示了来自图 6.30(a)的图像和它的补色图像。用于计算补色的RGB变换画在图6.33(b)中。它们等同于3.2.1 节定义的灰度变换。注意,计算补色使人想到通常照片的彩色负片。原图像的红色在补色中用青色代替,当原始图像是黑色时,补色是白的,等等。

补色图像中每种色调都可从原图像用图 6.32 的彩色图来预测。涉及补色计算的每一个RGB分量变换仅是对应的输入彩色分量的一个函数。不像图 6.31 的亮度变换,在这个例子中使用的RGB补色变换函数没有直接的HSI空间等效当量。作为一个作业留给读者(见习题 6.18),说明补色的饱和度分量不能单独从输入图像的饱和度分量计算出来。图 6.33(d)提供了一种使用图6.33(b)给出的色调、饱和度和亮度变换的补色的一种近似。注意,输入图像的饱和度分量是不可改变的,它对图6.33(c)和(d)之间的视觉差别有响应。

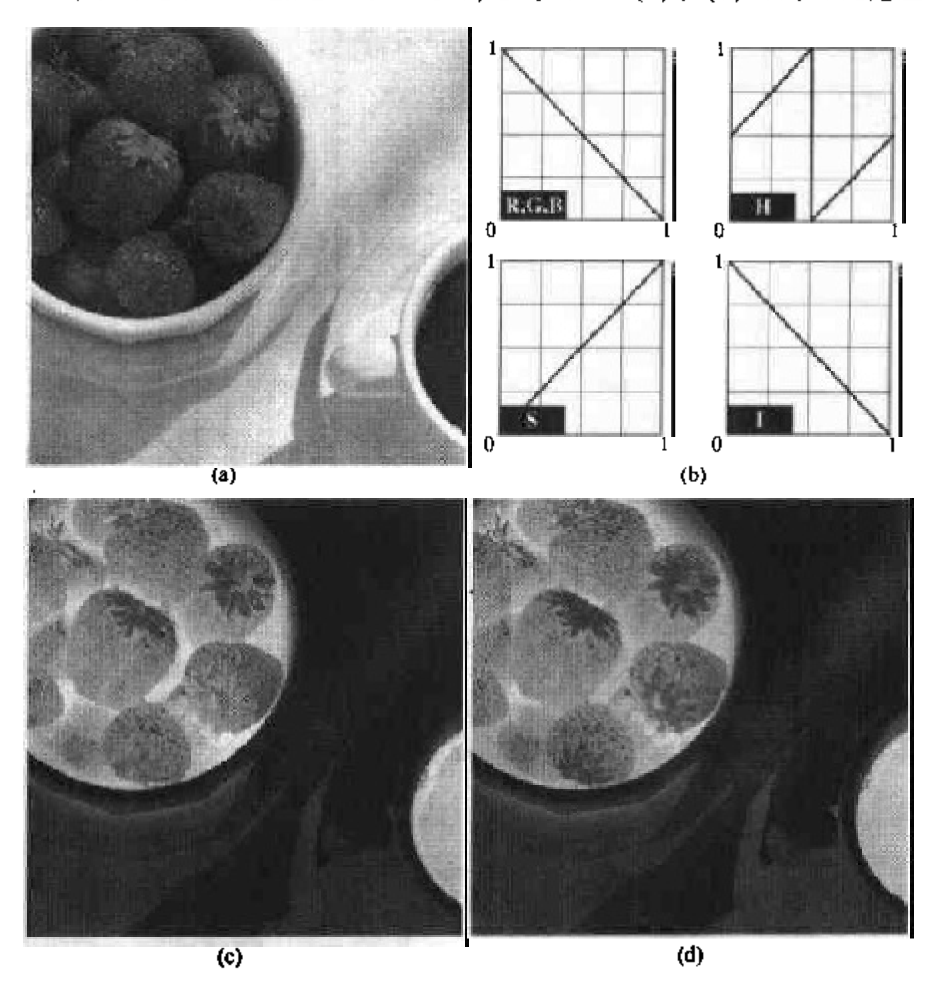


图 6.33 彩色分量变换。(a)原像,(b)补色变换函数,(c)基于 RGB 映射函数的(a)的补色,(d)用HSI变换的RGB补色的近似

6.5.3 彩色分层

突出图像中特殊的彩色区域对从其周围分离出目标物是很有用的。其基本思路是:(1)显示感兴趣的颜色以便从背景中把它们分离出来;(2)像模板那样使用由彩色定义的区域,以便进一步处理。最直接的方法沿用3.2.4节的灰度分层技术。然而,因为一个彩色像素是一个 n 维参量,彩色变换函数比图 3.11 相对应的灰度变换函数要复杂得多,事实上,所要求的变换比到目前为止考虑的彩色分量变换也复杂得多。这是因为所有的彩色分层方法都要求,每个像素变换后的彩色分量是所有 n 个原始像素彩色分量的函数。对一幅彩色图像分层的最简单的方法之一是,把某些感兴趣区域以外的区域的彩色映射为不突出的自然色,如果感兴趣的颜

色由宽为 W,中心在原彩色(平均)点并具有分量$(a_1, a_2, \cdots, a_n)$的立方体(或超立方体,$n>3$)所包围,则必要的变换集是:

$$s_i = \begin{cases} 0.5 & \left[\mid r_j - a_j \mid > \frac{W}{2} \right]_{\text{any}1 \leqslant j \leqslant n}, \\ r_i & \text{其他} \end{cases} \quad i = 1,2,\cdots,n \tag{6.5.7}$$

这些变换用强迫其他颜色为参考彩色空间的中点(任意选取的中性点)突出原型周围的颜色。例如,对于 RGB 彩色空间一个合适的自然点是灰度或彩色的中点(0.5,0.5,0.5)。

如果用一个圆球确定感兴趣的颜色,式(6.5.7)变为:

$$s_i = \begin{cases} 0.5 & \sum_{j=1}^{n} (r_j - a_j)^2 > R_0^2 \\ r_i & \text{其他} \end{cases} , i = 1,2,\cdots,n \tag{6.5.8}$$

这里 R_0 是封闭球形的半径(或超级球形,$n>3$),$(a_1, a_2, \cdots, a_n)$是其中心的分量(原彩色)。式(6.5.7)和式(6.5.8)另外有用的变化包括实现多彩色模型和在感兴趣区之外的区域减小彩色亮度,而不是赋以中性值。

例 6.8　彩色分层的一种说明

式(6.5.7)和式(6.5.8)可用于图 6.30(a)中,以分离出草莓食品。图6.34(a)和(b)显示使用两种变换的结果,在每种情况中,从最突出的草莓中选择具有RGB坐标(0.686 3,0.160 8,0.192 2)的原始红色;选择 W 和 R_0,则重点区域不必扩展到不希望的图像区域。实际值 $W=0.254\,9$, $R_0=0.176\,5$ 可用交互的方式确定。注意,式(6.5.8)基于球形的变换在包含更多的草莓红色区域的意义下稍好一些。半径为 0.176 5 的球形不完全包围宽度为 0.254 9的立方体,但它本身也不被立方体完全包围。

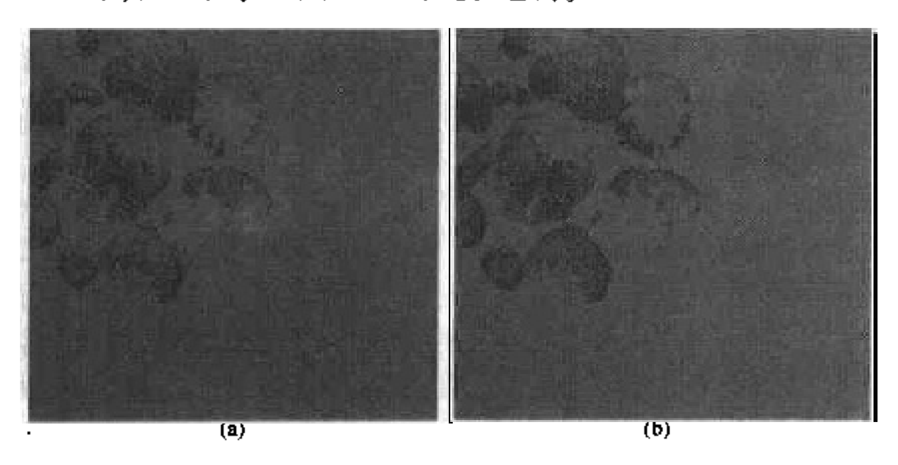


图 6.34　(a)在宽度为 $W=0.254\,9$,中心点在(0.686 3,0.160 8,0.192 2)的RGB立方体中检测红色的彩色分层变换,(b)在半径为 0.176 5,中心在相同点的RGB球形中检测红色,处在立方体和球形外边的像素由彩色(0.5,0.5,0.5)代替

6.5.4　色调和彩色校正

彩色变换可在多数台式计算机上进行。与数字摄像机、平面扫描仪及喷墨打印机相配合,就由个人计算机构成了数字暗室,该系统可以进行色调调整和彩色校正,高级彩色重现系统不

需要配备传统的湿式处理(暗室)设备就可进行这些处理。虽然色调和彩色校正在其他成像领域也很有用,但现在讨论的焦点集中在最通用的照片增强和彩色重现上。

在本节研究的变换效果归根到底是在打印中评价。因为此前研究、改进和评价的那些变换都是在监视器上,在所用的监视器和最终输出设备之间保持高度的彩色一致性是必要的。事实上,监视器上的彩色应该准确地显示任何数字扫描的原图像以及最后打印输出,这可用独立于设备的彩色模型实现,该模型与监视器的色移、输出设备以及其他设备有关。这一方法随用来把设备映射到模型及模型本身的彩色分布质量变化。许多彩色处理系统(CMS)选择的模型是CIE $L^*a^*b^*$ 模型,也称做CIELAB(CIE[1978], Robertson[1977])。$L^*a^*b^*$ 彩色分量由下式给出:

$$L^* = 116 \cdot h\left(\frac{Y}{Y_W}\right) - 16 \tag{6.5.9}$$

$$a^* = 500\left[h\left(\frac{X}{X_W}\right) - h\left(\frac{Y}{Y_W}\right)\right] \tag{6.5.10}$$

$$b^* = 200\left[h\left(\frac{Y}{Y_W}\right) - h\left(\frac{Z}{Z_W}\right)\right] \tag{6.5.11}$$

这里

$$h(q) = \begin{cases} \sqrt[3]{q} & q > 0.008\,856 \\ 7.787q + 16/116 & q \leq 0.008\,856 \end{cases} \tag{6.5.12}$$

X_W,Y_W 和 Z_W 是参考的白色三激励值——在CIE标准D65(图6.5的CIE色度图中,由 $x = 0.312\,7$, $y = 0.329\,0$ 定义)照明下典型的完美漫反射白色。$L^*a^*b^*$ 彩色空间是比色的(即与感觉色相匹配的编相同的码)、感觉一致的(即在各种色调中感觉彩色差别一致——见Mac Adams的经典论文[1942])和独立于设备的。虽然没有直接可显示的格式(需要变换到另外的彩色空间),其色域包括全部可见光谱并可以准确描述任何显示、打印或输入设备的彩色。类似 HSI系统,$L^*a^*b^*$ 系统是一个优秀的亮度和彩色分离器(L^* 表示亮度,a^* 表示红减绿,b^* 表示绿减蓝),使得它在图像操作(色调和对比度编辑)和图像压缩方面很有用。

标定图像系统的主要意义是,它允许对灰度和不平衡的彩色交互和独立地校正——它是按两个顺序操作,在诸如过饱和及欠饱和这样的彩色不规则问题解决之前,先校正图像灰度范围。一幅图像的灰度范围也叫做"主调型",它提供一般彩色强度的分布信息。高主调图像的多数信息集中在高亮度处,低主调图像的彩色主要位于低亮度处,中间主调图像位于其中间。正像在单色情况下,最好是彩色图像亮度在高亮度和阴影之间等同分布。下面的例子说明了对于灰度和彩色不平衡校正的各种彩色变换。

例6.9　灰度变换

改善图像色调的变换通常交互地选择。其概念是实验性地调整图像亮度和对比度以在合适的亮度上提供最大的细节,彩色本身并不改变。在RGB和CMY(K)空间,这意味着用相同的变换函数映射3(或4)个彩色分量;在HSI彩色空间,则仅改进亮度分量。

图6.35显示了典型的用于校正3个通常的灰度不平衡(即平淡的、亮的和暗的图像)的变换。在图6.35的第一行,S形曲线对提升对比度较理想[见图3.2(a)],它们中点固定,因此,高亮和阴影区可分别被加亮和变暗(该曲线反过来可用于校正对比度过大)。图中的

第二和第三行中的变换校正亮的和暗的图像，并使人想起图 3.6 的平方率变换。虽然彩色分量是离散的，如现在实际的变换函数，但它们本身以连续量显示和操作——典型情况是，从分段线性或高阶(对平滑的映射)多项式构造。注意，图 6.35 中图像的关键部分是直接可见的，它们还可以用图像彩色分量的直方图确定。

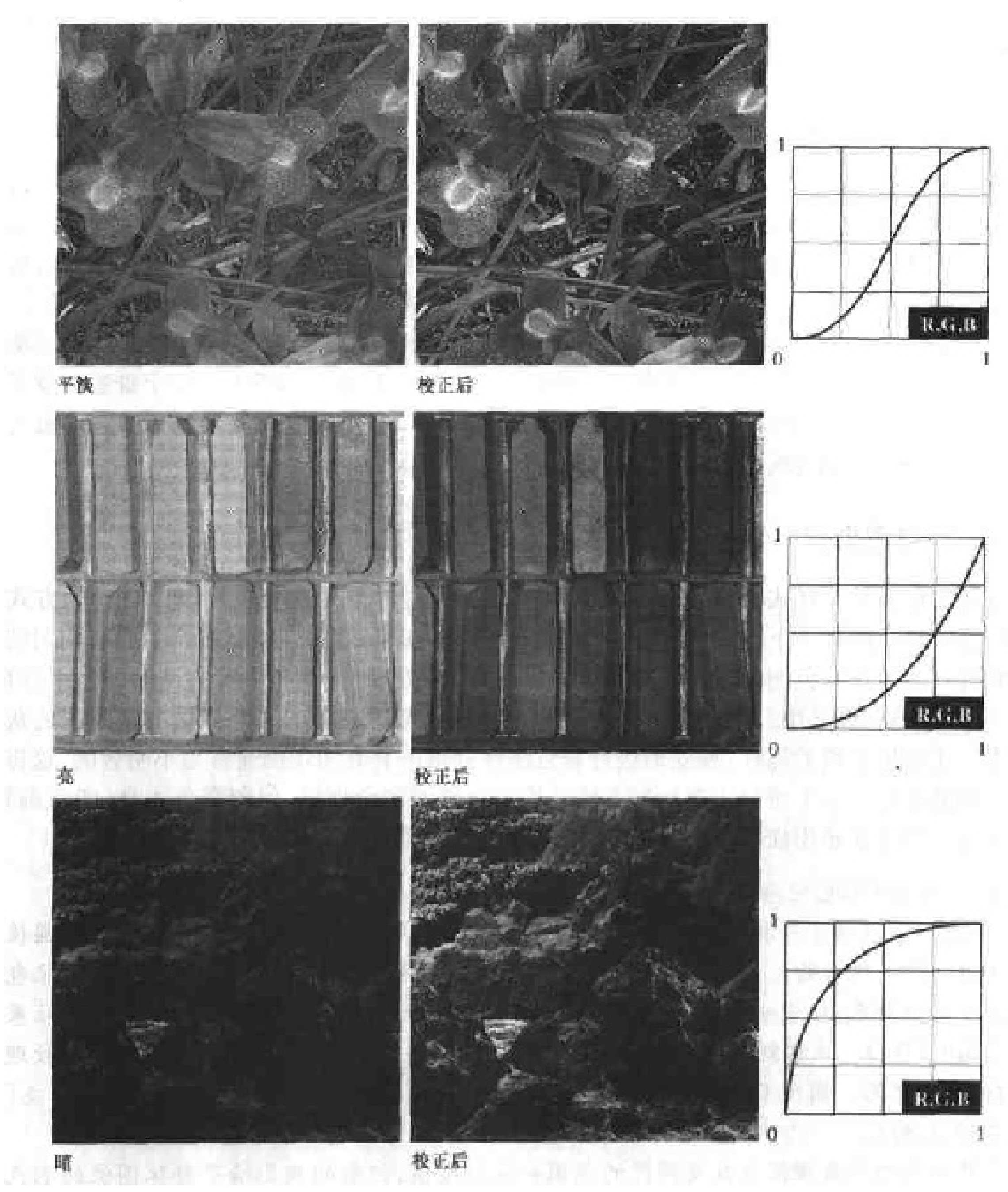


图 6.35　对平淡的、亮的(高主调)、暗的(低主调)彩色图像的校正。等量地调整红、绿、蓝分量没有改变图像色调

例 6.10 彩色平衡

在一幅图像的灰度特性已正确地确立了之后,就可着手进行彩色平衡的工作了。虽然彩色平衡可以客观地通过分析(用光谱仪)一幅图像中已知彩色来决定,当存在白色区域,即这里RGB或CMY(K)分量应当相等时,准确的视觉鉴定是可能的。正如在图6.36所看到的,对视觉彩色鉴别,皮肤色调是优秀的对象,因为人对一定的肤色高度敏感。鲜明的彩色,如亮红色物体,当做视觉彩色鉴别时没有什么价值。

当存在彩色不平衡时,有很多方法来校正它。在调整一幅图像的彩色分量时,完成的每一个步骤都影响图像的全部彩色平衡,这一点很重要。这就是说,对一种颜色的感觉受到周围彩色的影响。尽管图6.32的彩色轮可用于预测一个彩色分量如何影响其他彩色,例如,基于彩色轮,任何彩色的比例可用减小图像中相对色(或补色)总量的方式增加。类似地,可以提高两个直接邻接的彩色比例或减小两个邻接的彩色补色的百分比。例如,假定在RGB图像中有过多的深红色,则可以用(1)移去红和蓝,或(2)增加绿的方法减少它。

图6.36显示了用于校正简单的CMYK输出不平衡的变换。注意,变换是校正图像所要求的函数。使其反函数用来产生相关联的彩色不平衡。同时说明图像类似于暗室环境下的彩色印相,并且作为识别彩色打印问题的参考工具很有用。例如,过度的深红色导致太多的红色(左下图像)或太少的青色(如第二行最右边的图像)。

6.5.5 直方图处理

与前边章节的交互式增强方法不同,3.3节的灰度直方图处理变换可以以自动的方式作用于彩色图像。回忆一下,直方图均衡自动地确定一种变换,这种变换试图产生具有均匀的灰度值的直方图。在单色图像情况下,显示出(见图3.17)处理低、高和中等主调的图像很合理、成功,因为彩色图像是由多个分量组成的,所以,必须考虑适应多于一个分量的直方图的灰度级技术。正像所预料的那样,独立地进行彩色图像分量的直方图均衡通常是不明智的,这将产生不正确的彩色。一个更合乎逻辑的方法是均匀地扩展彩色强度,保留彩色本身(即色调)不变。下面的例子显示出HSI彩色空间是适合该种情况的理想方法。

例 6.11 在HSI彩色空间的直方图均衡

图6.37(a)显示了包括调味瓶和摇杯的调味瓶架台的彩色图像,其强度分量的范围值为[0,1](归一化后的)。正如在处理前的[图6.37(b)]强度分量直方图中看到的,图像包含大量的暗彩色,使中央强度减少到0.36。不改变色调和饱和度,均衡强度分量的结果示于图6.37(c)。注意到所有图像有效地加亮了,并且一些调味瓶和放调味瓶的木桌纹理现在都能看见。图6.37(b)显示了新图像的直方图以及用于均衡强度分量的变换[见式(3.3.8)]。

虽然强度均衡处理没有改变图像的色调和饱和度值,但它的确影响了整体图像的彩色感观。注意,(特别是不振动时)瓶中的油和醋。图6.37(d)显示了采用增加图像的饱和度分量,然后通过图6.37(b)的变换进行直方图均衡这一特殊校正的结果。这类调整在HSI空间处理强度分量时是很适用的。因为强度改变通常影响到图像的彩色状态。

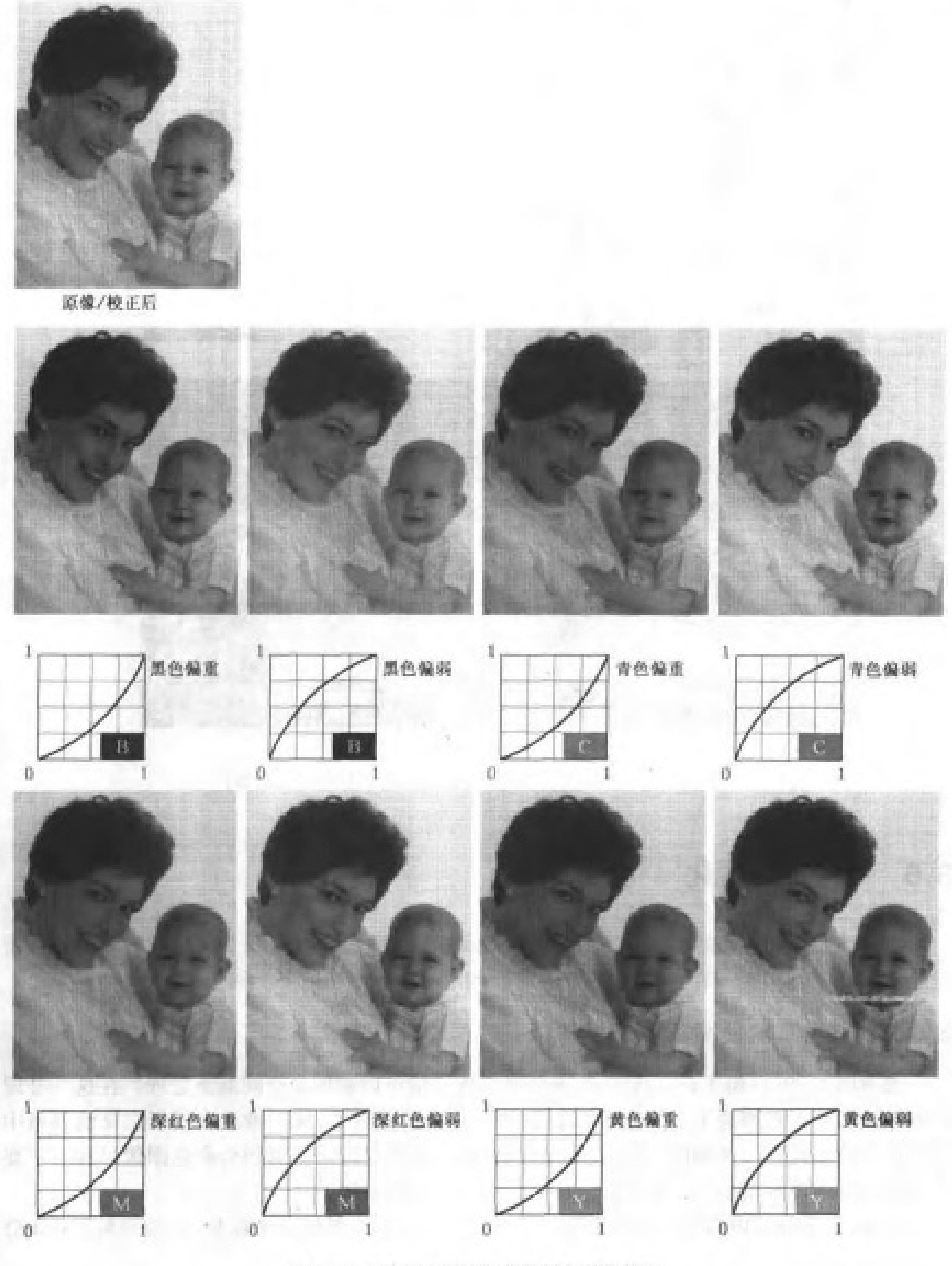


图 6.36　对 CMYK 图像进行彩色平衡校正

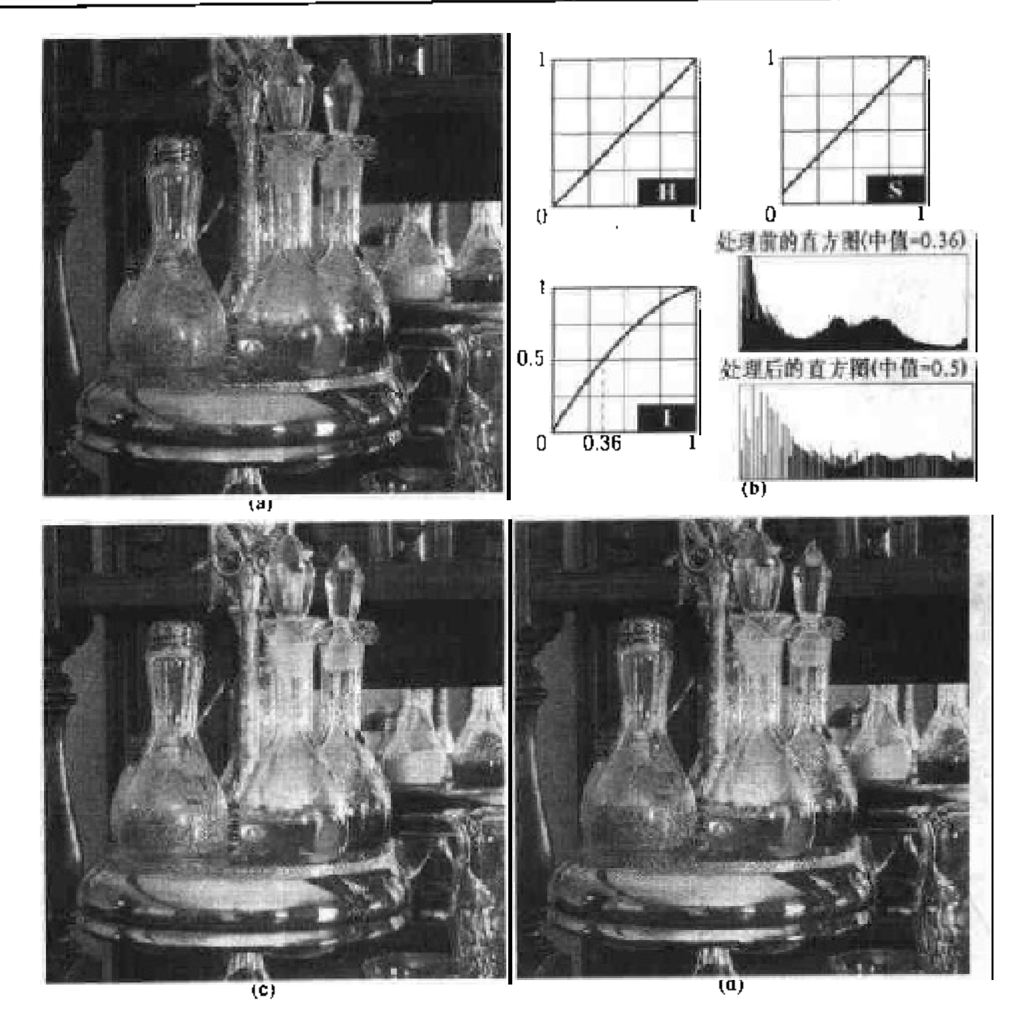


图 6.37 HIS 彩色空间的直方图均衡(后面是饱和度调整)

6.6 平滑和尖锐化

在不涉及相邻像素的情况下,对彩色图像每一个像素变换后的下一步是基于周围像素特点的改进处理。在这一节,这类邻域处理的基础示于彩色图像的平滑和尖锐化中。

6.6.1 彩色图像平滑

参考图 6.29(a)和 3.6 节的讨论,灰度级图像平滑可以看做是空间滤波处理。在这一处理中滤波模板的系数都是 1。当模板滑过图像时,图像被平滑了,每一像素由模板定义的邻域中像素的平均值代替。正如图6.29(b)所见到的,这一概念很容易扩展到全彩色图像处理。主要差别是代替灰度标量值,必须处理式(6.4.2)给出的分量向量。

令 S_{xy} 表示在RGB彩色图像中定义一个中心在(x,y)的邻域的坐标集,在该邻域中RGB分量的平均值为:

$$\bar{\mathbf{c}}(x,y)=\frac{1}{K}\sum_{(x,y)\in S_{xy}}\mathbf{c}(x,y) \tag{6.6.1}$$

它效仿式(6.4.2),向量的附加特性为:

$$\bar{\mathbf{c}}(x,y)=\begin{bmatrix}\frac{1}{K}\sum_{(x,y)\in S_{xy}}R(x,y)\\ \frac{1}{K}\sum_{(x,y)\in S_{xy}}G(x,y)\\ \frac{1}{K}\sum_{(x,y)\in S_{xy}}B(x,y)\end{bmatrix} \tag{6.6.2}$$

可以看出,如标量图像那样,该向量分量可以用传统的灰度邻域处理单独地平滑RGB图像的每一平面得到。

这样可以得出结论:用领域平均值平滑可以在每个彩色平面的基础上进行。其结果与用RGB彩色向量执行平均是相同的。

例6.12　用邻域平均进行彩色图像平滑

考虑显示于图6.38(a)中的彩色图像,在图6.38(b)到(d)描绘了该图像的红、绿、蓝3个平面。图6.39(a)到(c)显示图像的HSI分量。与前节讨论的一样,可以用3.6节的5×5的灰度平均模板平滑图6.38的RGB图像。简单地单独平滑RGB彩色平面的每一个,然后混合处理过的平面,以形成一幅平滑的全彩色结果。这样计算过的图像示于图6.40(a)。注意,它的显示正如在3.6节讨论和实验所希望的那样。

在6.2节指明HSI彩色模型的重要优点是解除了强度和彩色信息的关系(与灰度关系密切)。这就使得许多灰度处理技术适合彩色处理,并且可能仅对图6.39中HSI描述的强度分量平滑更有意义。为了说明这一方法的优点和重要性,下边仅对强度分量进行平滑(保留色调和饱和度分量不变)并把处理结果变换为RGB图像加以显示。平滑后的图像显示于图6.40(b)。注意到它与图6.40(a)很相似,但是正如从图6.40(c)不同图像看到的那样是有差别的。这是由于两个不同颜色的像素平均是两种彩色的混合,而不是原色混合。仅对强度平滑,图6.40(b)的图像保留了它的原色调和饱和度——即保留了它的原彩色。最后,注意到,其差别(在这个例子中是平滑后的结果)将随着平滑模板的增加而增加。

6.6.2 彩色图像尖锐化

在这一节考虑用拉普拉斯方法的图像尖锐化处理(见3.7.2节)。从向量分析知道向量的拉普拉斯被定义为一向量,其分量等于输入向量的独立标量分量的拉普拉斯微分。

在RGB彩色系统中,式(6.4.2)中的向量$\mathbf{c}$的拉普拉斯变换为:

$$\nabla^2[\mathbf{c}(x,y)]=\begin{bmatrix}\nabla^2R(x,y)\\ \nabla^2G(x,y)\\ \nabla^2B(x,y)\end{bmatrix} \tag{6.6.3}$$

正如前节讲述的,它告诉我们可以通过分别计算每一分量图像的拉普拉斯去计算全彩色图像的拉普拉斯。



图 6.38 (a)RGB 图像,(b)红分量图像,(c)绿分量图像,(d)蓝分量图像

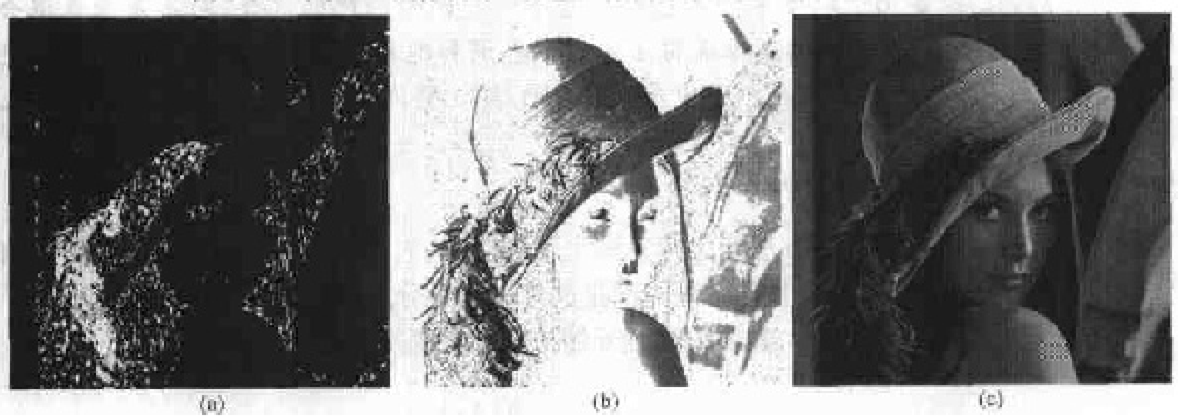


图 6.39 图 6.38 中 RGB 彩色图像的 HSI 分量。(a)色调,(b)饱和度,(c)强度

例 6.13 用拉普拉斯尖锐化

图 6.41(a)是通过用式(3.7.6)计算图 6.38 中 RGB 分量图像的拉普拉斯,并混合它们产生尖锐化的全彩色结果产生的。图6.41(b)显示了基于图 6.39 的HSI分量相似的尖锐化结果。这个结果是混合强度分量的拉普拉斯而色调和饱和度分量不变的结果。RGB和基于HSI的结果间的差别示于图6.41(c),它源于例 6.12 解释的相同原因。



图 6.40　用 5×5 平均模板平滑图像。(a)处理 RGB 每一分量图像的结果,(b)处理HSI图像强度分量,转换为RGB图像的结果,(c)两种结果间的差别

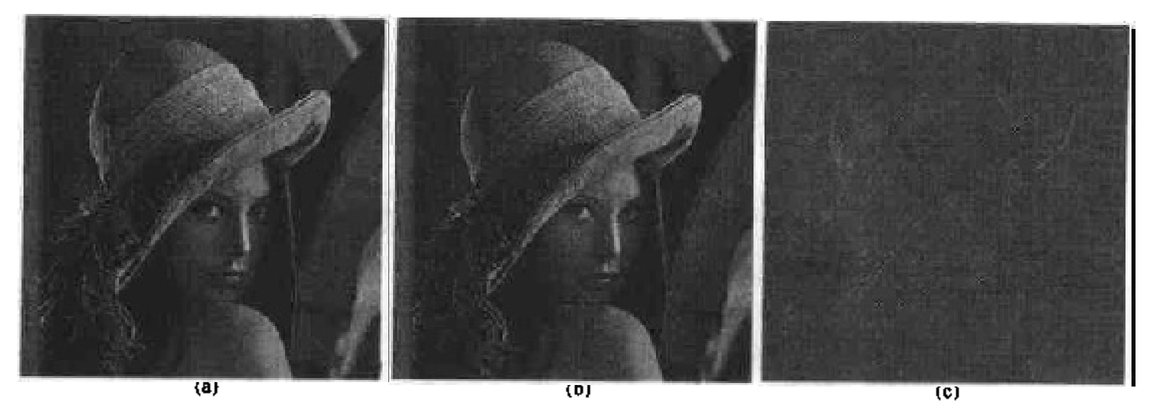


图 6.41　用拉普拉斯进行图像尖锐化,(a)处理每一个RGB通道的结果,(b)处理强度分量并转换为 RGB 的结果,(c)两种结果的差别

6.7　彩色分割

分割是把一幅图像分成区域的处理。虽然分割是第 10 章讨论的话题,这里出于连续性的原因,简单地介绍一下彩色分割。在下面的讨论中,读者不会有困难。

6.7.1　HSI 彩色空间分割

如果希望基于彩色分割一幅图像,并且想在单独的平面上执行处理,会很自然地首先想到HSI空间。因为在色调图像中描述彩色是很方便的。典型情况是,为了在色调图像中分离出感兴趣的特征区,饱和度被用做一个模板图像。在彩色图像分割中强度图像不常使用,因为它不携带彩色信息。下面的例子是典型的如何在HSI系统中执行分割的例子。

例 6.14　在 HSI 空间的分割

假定兴趣是分割图 6.42(a)中左下角的图像中的微红色区,虽然它是用伪彩色产生的,但是可以像全彩色图像那样处理而通常不会有损失。图6.42(b)到(d)是它的HSI分量图像。注意,比较图6.42(a)和(b),我们感兴趣的区域有相对高的色调值,表明彩色在红色的蓝及深红一侧(见图 6.13)。图6.42(e)显示了门限产生的二值图像,该图像在饱和度图像中门限值等于最大饱和度的 10%,任何比门限大的像素值赋 1 值(白),其他赋 0 值(黑)。

图 6.42(f)用色调图像产生模板,图 6.42(g)是结果图像的直方图(注意,灰度尺度在[0,1]范围内)。在直方图中我们看到高值(这是感兴趣的值)被分组到灰度标尺的最高端,接近 1.0。具有 0.9 的门限值的门限化图像导致的二值图像示于图6.42(h)。在该图像中白点的空间位置识别原像点,原像点具有感兴趣的微红色调。这距完善的分割相差甚远,因为在原图像中还有一些应该有微红色调的点,但没有被该分割方法识别出来。然而实验表明图 6.42(h)所示的白色区在识别原图像的微红色分量时,是这种方法能做到的最好结果。在下面一节讨论的分割方法可产生相当好的结果。

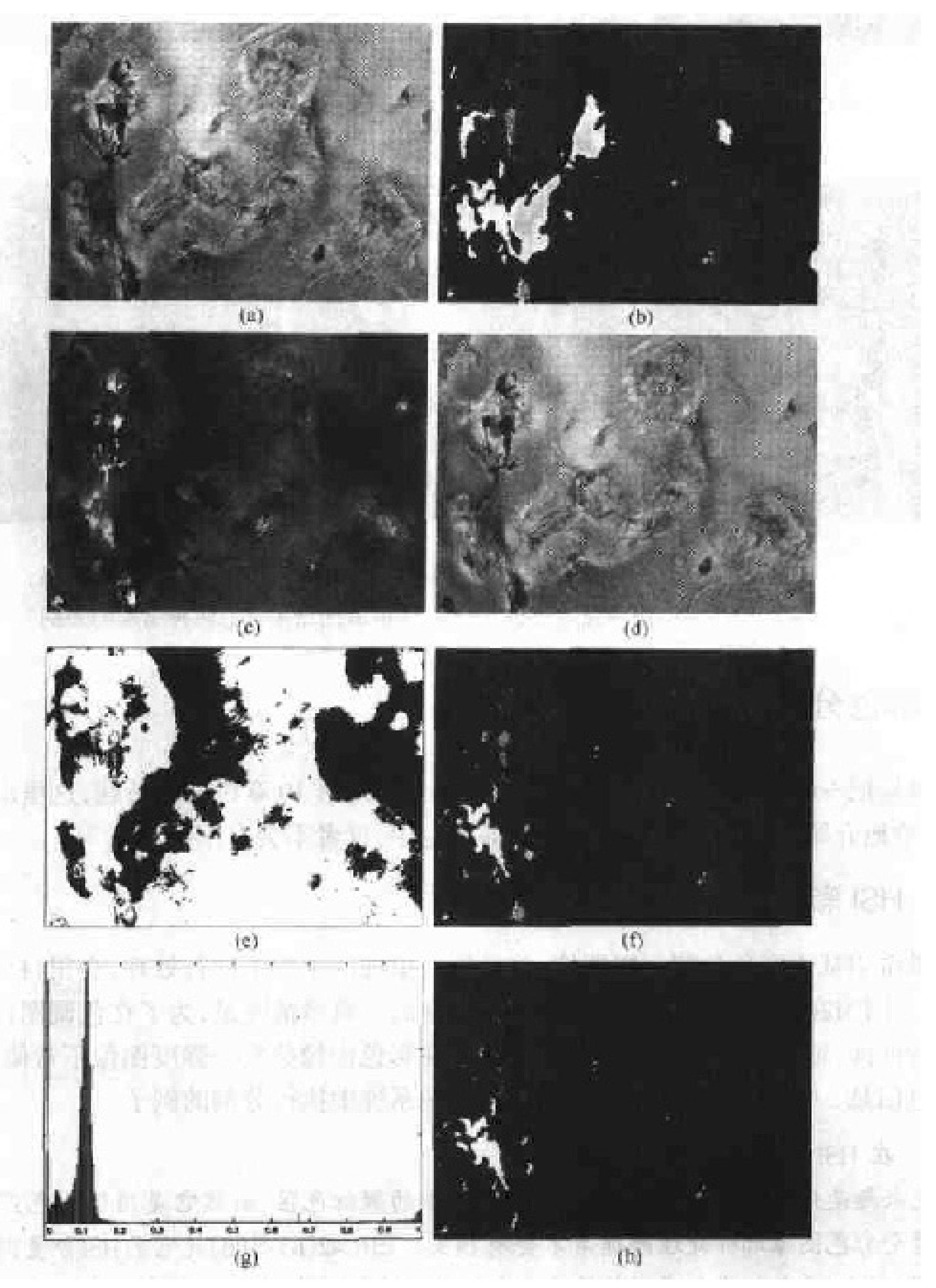


图 6.42 在 HSI 空间的图像分割。(a)原像,(b)色调,(c)饱和度,(d)强度,(e)二值饱和度模板(黑 = 0),(f)(b)和(e)相乘,(g)(f)的直方图,(h)(a)中红分量的分割

6.7.2 RGB 向量空间分割

正如在这一节多次提到的，虽然在 HSI 空间的工作更直观，但分割是这样一个领域，即通常用RGB彩色向量得到的结果更好。该方法是直接的。假如目标是在RGB图像中分割特殊彩色区域的物体。给定一个感兴趣彩色的有代表性的彩色点样品集，可得到一个彩色"平均"估计，这种彩色是我们希望分割的彩色。令这个平均彩色用RGB向量 $\mathbf{a}$ 来表示。分割的目标是对给定图像中每一个RGB像素进行分类，因为在确定的范围会有某种颜色或者没有这种颜色。为了执行这一比较，有一个相似性度量是必要的。最简单的度量之一是欧氏距离。令 $\mathbf{z}$ 代表RGB空间中的任意一点，如果它们之间的距离小于特定的阈值 D_0，我们说 $\mathbf{z}$ 与 $\mathbf{a}$ 是相似的，$\mathbf{z}$ 和 $\mathbf{a}$ 间的欧氏距离由下式给出：

$$\begin{aligned} D(\mathbf{z},\mathbf{a}) &= \| \mathbf{z} - \mathbf{a} \| \\ &= \left[(\mathbf{z} - \mathbf{a})^T (\mathbf{z} - \mathbf{a}) \right]^{\frac{1}{2}} \\ &= \left[(z_R - a_R)^2 + (z_G - a_G)^2 + (z_B - a_B)^2 \right]^{\frac{1}{2}} \end{aligned} \tag{6.7.1}$$

这里注脚 R,G,B 表示向量 $\mathbf{a}$ 和 $\mathbf{z}$ 的RGB分量。$D(\mathbf{z},\mathbf{a}) \leqslant D_0$ 的点的轨道是半径为 D_0 的实心球，如图6.43(a)所示。包含在球内部和表面上的点符合特定的彩色准则；球外面上的点则不符合准则。在图像中对这两类点集编码，比如说黑或白，产生一幅二值分割图像。

式(6.7.1)一个有用的推广是如下形式的距离测度：

$$D(\mathbf{z},\mathbf{a}) = \left[(\mathbf{z} - \mathbf{a})^T \mathbf{C}^{-1} (\mathbf{z} - \mathbf{a}) \right]^{\frac{1}{2}} \tag{6.7.2}$$

这里 $\mathbf{C}$ 是希望分割的彩色典型的样本协方差矩阵[①]。$D(\mathbf{z},\mathbf{a}) \leqslant D_0$ 的点的轨道描述了一个实心的三维球体[图 6.43(b)]，其具有的最大特点是，主轴在最大数据范围的方向旋转。当 $\mathbf{C} = \mathbf{I}$ 时，则 3×3 单位矩阵[式(6.7.2)]简化为式(6.7.1)。分割如前一节描述的那样。

因为距离是正的和单调的，可以用距离的平方运算来代替，这样就避免了开方运算。然而，即使不计算平方根，执行式(6.7.1)或式(6.7.2)的计算代价也很高。折中方案是使用边界盒，如图6.43(c)所示。在该方法中，盒的中心在 $\mathbf{a}$ 上，沿每一彩色轴的尺度选择与沿每一个轴取样的标准差成比例，标准差的计算只使用一次样本彩色数据。

给定一个任意的彩色点，根据它是否在盒子表面或内部来进行分割，如同用距离方法一样。然而，判定一个彩色点是在盒子内部还是外部的计算比求球或椭球界限时要简单得多。注意，前边的讨论是6.5.3节介绍的彩色分割方法的扩展。

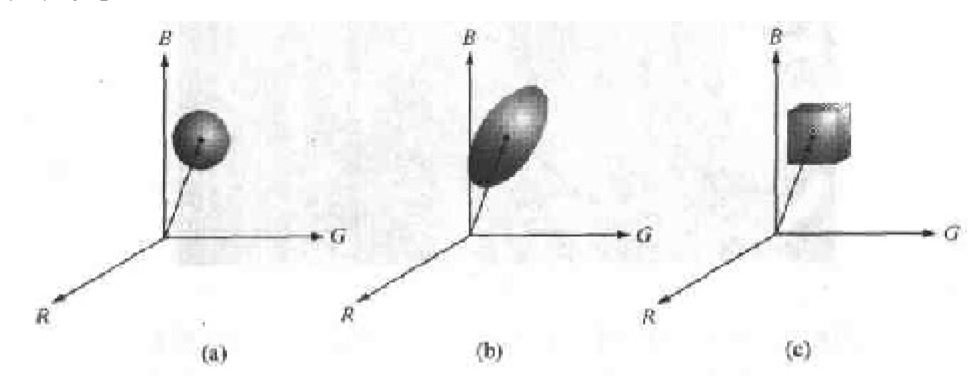


图 6.43　对于 RGB 向量分割，封闭数据范围的 3 种方法

① 一组向量样本协方差矩阵计算的详细讨论见 11.4 节。

例 6.15 在 RGB 空间的彩色分割

图 6.44(a)所示的方形区域包含微红色样品,这里希望分割出彩色图像。这是在例 6.14 中利用色调考虑的相同问题。但是在这里,研究该问题时利用RGB彩色向量。后面的方法利用包含在图6.44(a)中长方形中的彩色点计算平均向量 **a**,然后计算这些样点的红,绿,蓝的标准差。盒子的中点在 **a**,它的尺度沿每一个RGB轴以沿相应轴的数据标准差的 1.25 倍选择。例如,令 σ_R 代表样点红分量的标准差,然后,盒子的尺度沿红轴从$(a_R-1.25\sigma_R)$扩展到$(a_R+1.25\sigma_R)$,这里 a_R 代表平均向量 **a** 的红分量。在整个彩色图像中编码每一点的结果为:如果点位于盒子表面或内部为白色,否则为黑色,如图6.44(b)所示。注意,分割的区域如何从矩形封闭的彩色样本推广。事实上,比较图 6.44(b) 和图 6.42(h),我们看到在RGB向量空间分割会产生准确得多的结果。在一定意义上,在原彩色图像中它们相对于我们定义的"微红色"更接近。

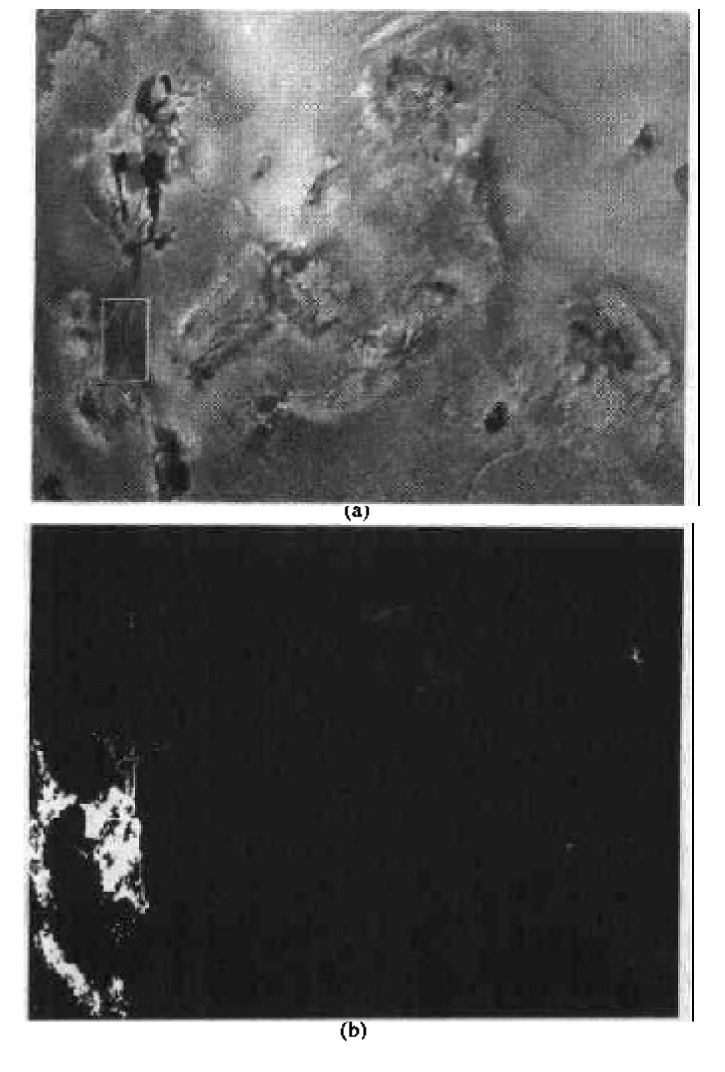


图 6.44 RGB 空间分割。(a)由封闭的矩形所示的感兴趣的彩色原像,(b)与图6.42(h)相比,RGB向量空间分割的结果

6.7.3　彩色边缘检测

正如第10章讨论的那样，边缘检测对图像分割是一个重要的工具。在这一节，我们有兴趣比较一下以单独图像为基础计算边缘和在彩色空间直接计算边缘的问题。基于边缘的分割的详细内容由10.1.3节给出。

梯度算子边缘检测在3.7.3节关于边缘增强的部分中已介绍过了。遗憾的是，3.7.3节中讨论的梯度对向量没有定义。这样，立刻会想到分别计算图像梯度，然后形成彩色图像将会导致错误的结果。一个简单的例子可帮助我们说明其原因。

考虑两幅 $M\times M$ 彩色图像（M 为奇数）[图6.45(d)和(h)]，它们分别由图6.45(a)到(c)和图6.45(e)到(g)中的3个分量图像合成。例如，如果计算每一个分量图像的梯度[见式(3.7.13)]，并将结果相加形成两幅相应的RGB梯度图像，则在点$[(M+1)/2,(M+1)/2]$处梯度值在两种情况下都将相同。直观地看，我们希望图6.45(d)中图像那一点的梯度更强，因为R,G,B图像的边缘在该图像中处在相同的方向上，与图6.45(h)的图像相反，其只有两个边缘在相同的方向上。从这个简单的例子可以看到处理3个独立平面形成的合成梯度图像可导致错误结果。如果问题是只检测边缘中的一个，则单独分量方法常可得到可接受的结果。然而，如果准确度是重点，则很明显需要一个可用于向量的梯度的新定义。下面讨论Di Zenzo[1986]为此提出的一种方法。

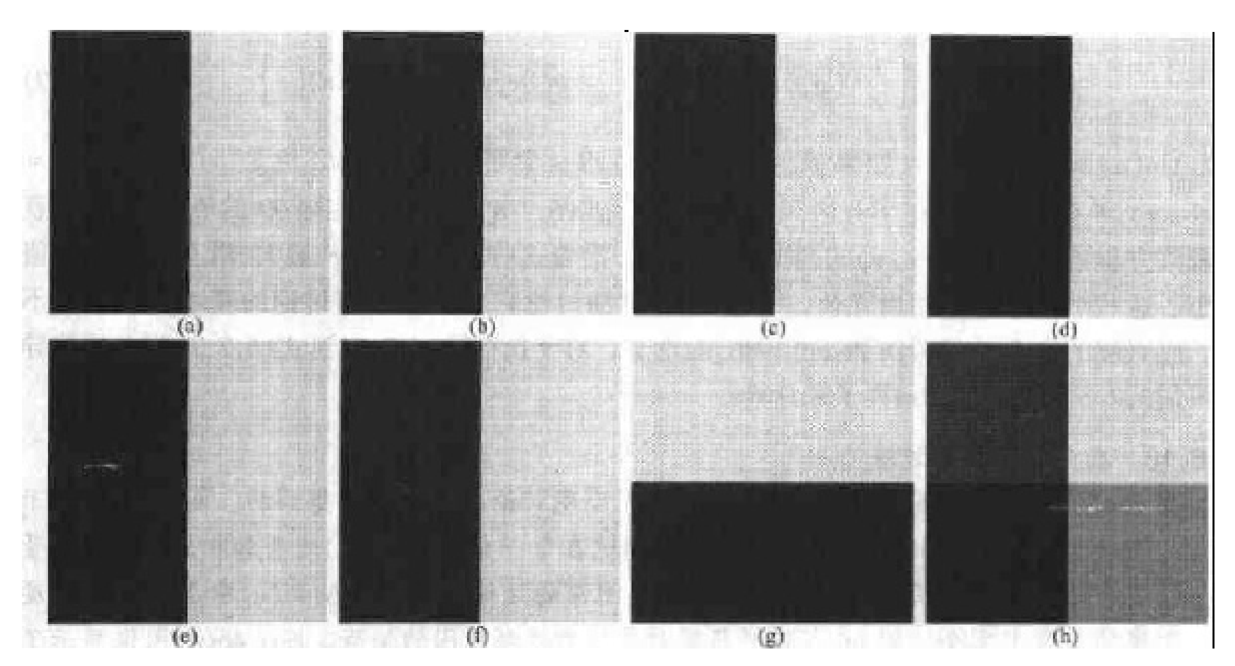


图6.45　(a)~(c) R,G,B 分量图像，(d)产生的彩色图像，(f)~(g)R,G,B分量图像，(h)产生的彩色图像

首要的问题是定义式(6.4.2)中向量 $\mathbf{c}$ 在任意点(x,y)处的梯度（幅度和方向）。正如刚刚提到的，在3.7.3节研究的梯度可用于标量函数 $f(x,y)$，但不能用于向量函数。下面是各种方法中的一种，在这种方法中可以针对向量扩展梯度的概念。回忆一下标量函数 $f(x,y)$，梯度是在坐标(x,y)处指向 f 的最大变化率方向的向量。

令 $\mathbf{r}$,$\mathbf{g}$ 和 $\mathbf{b}$ 是RGB彩色空间(图 6.7)沿 R,G,B 轴的单位向量,可定义向量为:

$$\mathbf{u} = \frac{\partial R}{\partial x}\mathbf{r} + \frac{\partial G}{\partial x}\mathbf{g} + \frac{\partial B}{\partial x}\mathbf{b} \tag{6.7.3}$$

和

$$\mathbf{v} = \frac{\partial R}{\partial y}\mathbf{r} + \frac{\partial G}{\partial y}\mathbf{g} + \frac{\partial B}{\partial y}\mathbf{b} \tag{6.7.4}$$

数量 g_{xx},g_{yy}和g_{xy}定义为这些向量的点乘,如下所示:

$$g_{xx} = \mathbf{u}\cdot\mathbf{u} = \mathbf{u}^T\mathbf{u} = \left|\frac{\partial R}{\partial x}\right|^2 + \left|\frac{\partial G}{\partial x}\right|^2 + \left|\frac{\partial B}{\partial x}\right|^2 \tag{6.7.5}$$

$$g_{yy} = \mathbf{v}\cdot\mathbf{v} = \mathbf{v}^T\mathbf{v} = \left|\frac{\partial R}{\partial y}\right|^2 + \left|\frac{\partial G}{\partial y}\right|^2 + \left|\frac{\partial B}{\partial y}\right|^2 \tag{6.7.6}$$

$$g_{xy} = \mathbf{u}\cdot\mathbf{v} = \mathbf{u}^T\mathbf{v} = \frac{\partial R}{\partial x}\frac{\partial R}{\partial y} + \frac{\partial G}{\partial x}\frac{\partial G}{\partial y} + \frac{\partial B}{\partial x}\frac{\partial B}{\partial y} \tag{6.7.7}$$

R,G 和 B 及由此而来的 g 是 x 和 y 的函数。利用该表示法,$\mathbf{c}(x,y)$的最大变化率方向可以由角度给出(Di Zenzo[1986]):

$$\theta = \frac{1}{2}\arctan\left[\frac{2g_{xy}}{(g_{xx} - g_{yy})}\right] \tag{6.7.8}$$

(x,y)点在 θ 方向上变化率的值由下式给出:

$$F(\theta) = \left\{\frac{1}{2}[(g_{xx} + g_{yy}) + (g_{xx} - g_{yy})\cos 2\theta + 2g_{xy}\sin 2\theta]\right\}^{\frac{1}{2}} \tag{6.7.9}$$

因为 $\tan(\alpha) = \tan(\alpha \pm \pi)$,如果 θ_0 是式(6.7.8)的一个解,则 $\theta_0 \pm \pi/2$ 也是。因此 $F(\theta) = F(\theta + \pi)$,$F$ 仅需对θ 值在半开区间$[0,\pi)$计算。式(6.7.8)提供两个相隔 90°的值,这一事实意味着该方程涉及每一点(x,y)的两个正交方向。沿着这些方向之一,F 最大,沿其他方向其值最小。这些结果的推导相当冗长,并且在这里详细讨论它对我们当前讨论的基本目标收益不大。感兴趣的读者可参考Di Zenzo[1986]的论文。对于执行式(6.7.5)到式(6.7.7)要求的偏导数可用3.7.3节讨论的Sobel算子来计算。

例 6.16　在向量空间的边缘检测

图 6.46(b)是图 6.46(a)图像的梯度,它是用刚刚讨论的向量方法得到的。图6.46(c)显示了由计算每一个RGB分量图像的梯度,并通过在每一坐标点(x,y)处叠加相应的 3 个分量值混合的梯度图像获得的图像。向量梯度图像的边缘细节比图6.46(c)中单独平面梯度图像的细节更完全。例如,可观察研究对象的右眼部周围的细节。图6.46(d)图像显示了在每一点(x,y)处的两种梯度图像间的差别。注意一下两种方法得到的可接受的结果。图6.46(b)中额外的细节是否值得附加计算量(与Sober算子的执行相反,它用于产生单独平面的梯度)仅由给定问题的要求决定。图6.47显示了 3 个分量梯度图像,在相加和标定时可用来得到图6.46(c)。

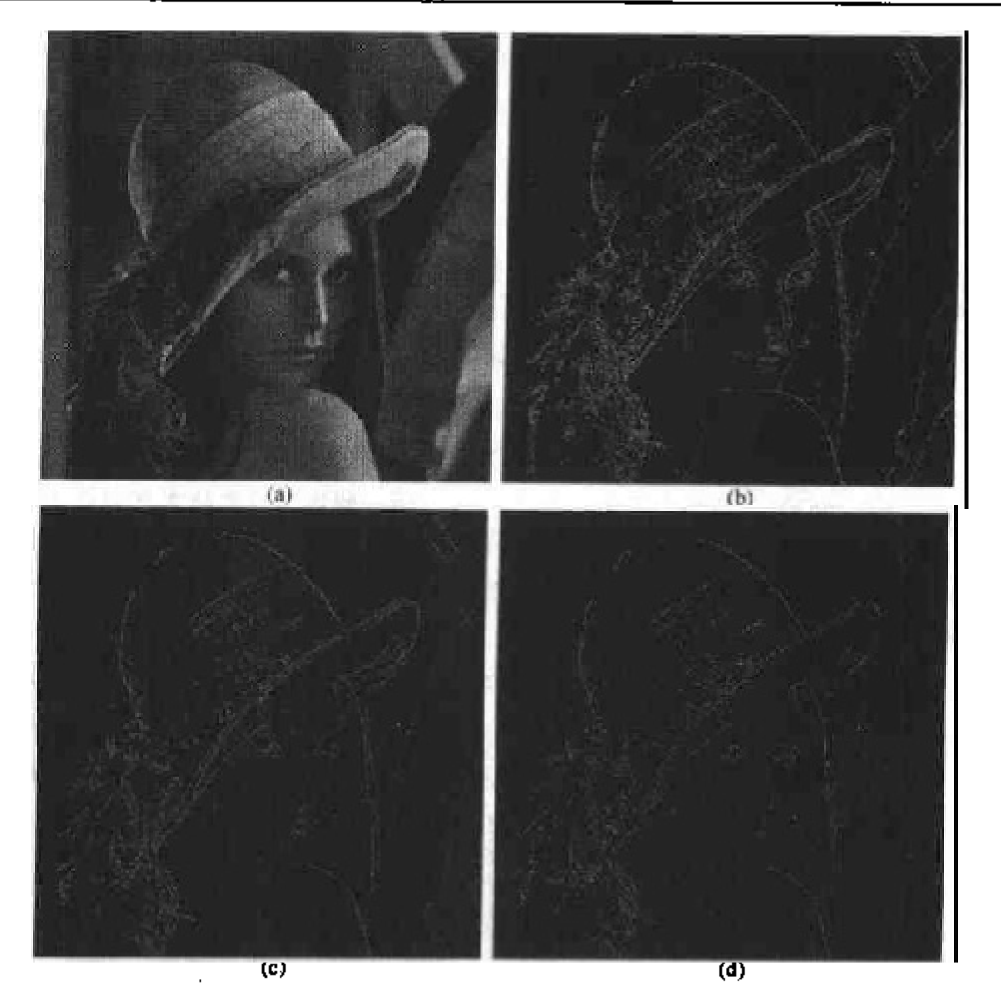


图 6.46　(a)RGB 图像,(b)在 RGB 彩色空间计算的梯度,(c)在分量图像基础上计算然后相加的梯度,(d)(b)和(c)间的差别

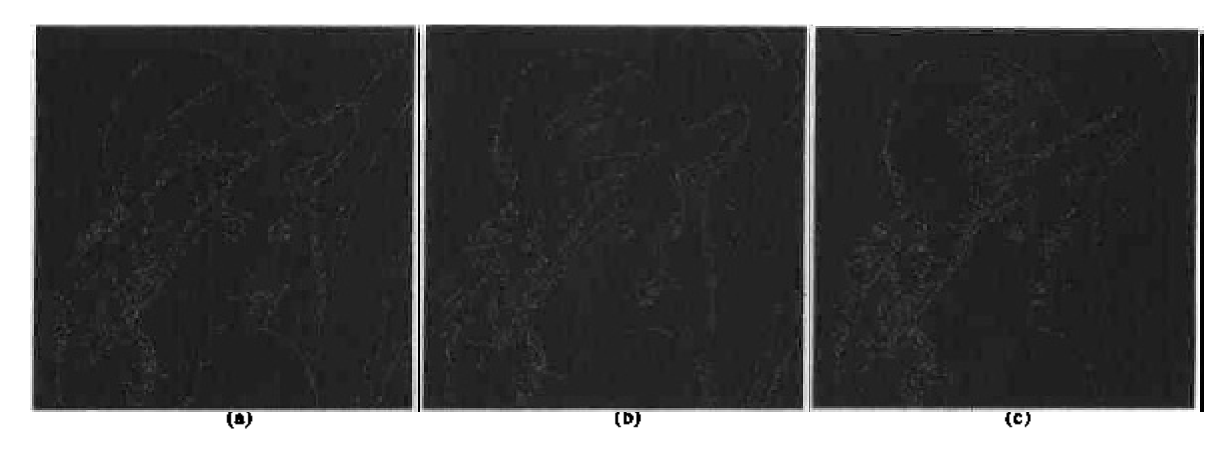


图 6.47　图 6.46 彩色图像的分量梯度图像。(a)红分量,(b)绿分量,(c)蓝分量。这3幅图像相加并标定产生的图像

6.8 彩色图像的噪声

5.2 节讨论的噪声模型可用于彩色图像。通常,彩色图像噪声的内容在每个彩色通道中具有相同的特性,但是,可能对不同彩色通道造成的影响不同。一种可能是个别通道的电子学故障。然而,不同的噪声程度更像是由用于每个彩色通道的相对照射强度的差异造成的。例如,CCD摄像机使用红滤色器将减小用于红传感器的照射强度。CCD传感器在低照明水平下就是噪声源,因此,在这种情况下,导致RGB图像红色分量比其他两个分量往往容易引入噪声。

例 6.17 把 RGB 图像噪声转换为 HSI 的效果说明

在这个例子中,简单看一下彩色图像中的噪声及当从一个彩色模型转换为另一个模型时如何转移噪声。图6.48(a)到(c)显示了由高斯噪声退化的一幅RGB图像的 3 个彩色平面,图6.48(d)是合成的RGB图像。注意,细网纹噪声在彩色图像中相比单色图像不太容易引起注意。图 6.49(a)到(c)显示把图6.48(d)中的RGB图像转换为HSI图像的结果。用原图像的HSI分量(图 6.39)比较这些结果,并注意如何有效地减小噪声图像色调与饱和度分量。这分别是由式(6.2.2)和式(6.2.3)求余弦与取最小值操作的非线性造成的。另一方面,图6.49(c)中的强度分量比 3 个有噪声的RGB分量图像中的任一个都稍微平滑一些。这是由于强度图像是RGB图像的平均,如式(6.2.4)所指出的那样(回忆3.4.2节关于对图像求平均减小随机噪声的讨论)。

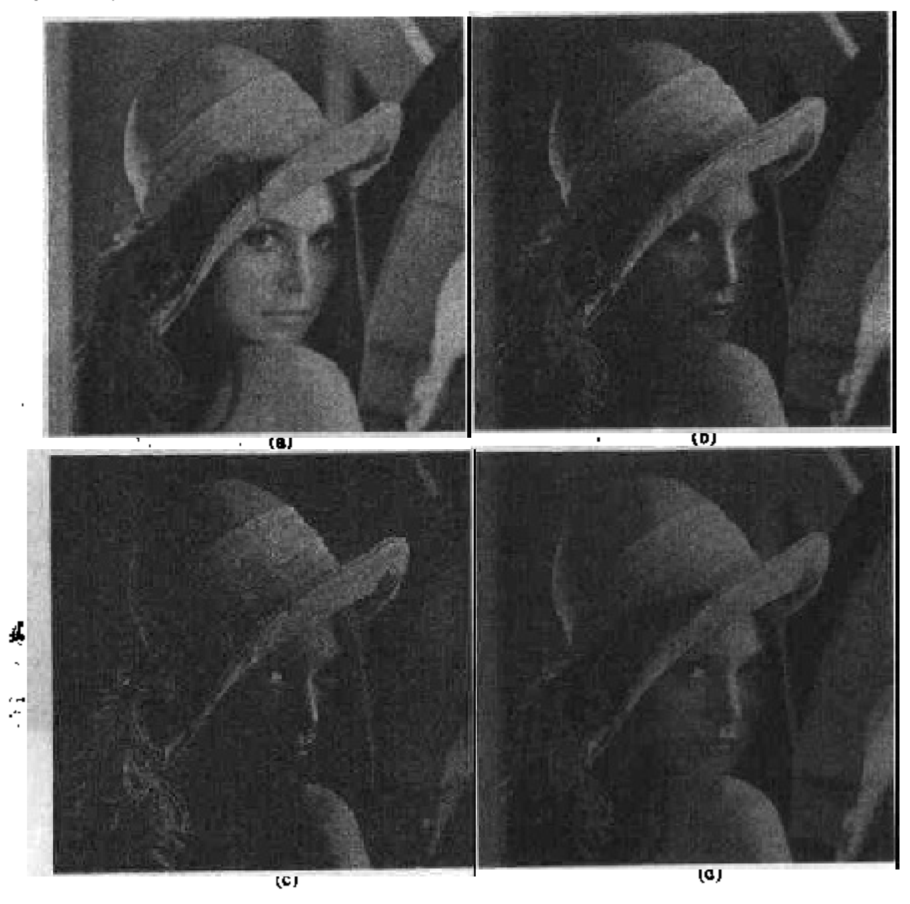


图 6.48 (a)~(c)由 0 均值、方差 800 的高斯噪声污染的红、绿和蓝分量图像,(d)RGB 结果图像[与图 6.46(a)比较]

在这种情况下，我们说，仅在RGB通道受噪声影响并转换到HSI空间后才把噪声散布到所有的HSI分量图像上了。图6.50(a)显示了一个例子。图6.50(a)显示了一幅RGB图像，其中绿图像被椒盐噪声污染，椒盐噪声的概率为0.05，图6.50(b)到(d)的HSI分量图像清楚地显示噪声怎样从绿的RGB通道散布到所有的HSI图像上。当然，这是我们不希望的，因为HSI分量的计算用到RGB所有的分量，正如6.2.3节所示那样。



图6.49 图6.48(d)含噪声的彩色图像的HSI分量。(a)色调，(b)饱和度，(c)强度



图6.50 (a)绿平面受椒盐噪声污染的RGB图像，(b)HSI图像的色调图像，(c)饱和度分量，(d)强度分量

迄今为止讨论过的处理都是正确的,全彩色图像滤波可以以图像为基础或依靠该过程直接在彩色空间进行。例如,用均值滤波减小噪声是6.6.1节讨论过的方法,在彩色空间,当对分量图像单独进行类似的处理时,我们知道它可以给出相同的结果。然而,其他滤波器却不能按这样公式化。这些例子包括在5.3.2节讨论过的各类统计滤波器。例如,在彩色向量空间执行中值滤波就必须找到一种向量排序方案,以便确定中值。当处理标量时这是一个简单的过程。当处理向量时,这一过程就相当复杂。向量排序的问题在这里超出了讨论的范围,但是,由Plataniotis和Venetsanopoulos[2000]所著的书,在向量排序和某些基于排序概念的滤波器方面是一本很好的参考书。

6.9 彩色图像压缩

因为描述彩色要求的比特数比描述灰度的数量大3到4倍,所以数据压缩在存储和传输彩色图像中起核心作用。对于前边章节的RGB、CMY(K)和HSI图像,作为压缩对象的数据是各彩色像素的分量(即,在RGB图像中像素的红,绿,蓝分量);它们是传递彩色信息的方法。压缩是减小或消除冗余度或者不相干数据的处理。虽然压缩是第8章的话题,但我们可以利用彩色图像在后边的例子中简要地说明一下这一概念。

例6.18 彩色图像压缩实例

图6.51(a)显示了一幅蝴蝶花的24比特RGB全彩色图像。图像中表示红、绿、蓝分量时都使用了8比特。图6.51(b)是从(a)中图像的一种压缩方案中重建的图像,事实上该图像是压缩然后解压缩后该图像的一个近似。虽然压缩后的图像不能直接显示,在输入到彩色监视器之前必须解压缩,但压缩图像对原图像的每230比特数据仅包含1数据比特(即1个存储比特)。假定压缩后的图像可通过Internet传输1分钟,那么原图像的传输至少要4小时。当然,传输后的数据必须解压缩后才能观看,但解压缩在几秒钟就可以做完。用于产生图6.51(b)的JPEG 2000压缩算法是当前引入的标准,这个标准在8.6.2节叙述。注意,重建的近似图像有点模糊。这是许多有损压缩技术的特点;可以通过改变压缩程度加以减小或消除。

小结

这一章的材料介绍了彩色图像处理,所涵盖的主题就图像处理这一分支所用的技术给读者提供了一个整体背景。彩色基本原理和彩色模型的论述是作为基础材料提供的,这些材料在技术规范和应用领域方面都在合适的范围内。特别地,把焦点集中在了彩色模型上,我们感觉到,它们不仅在数字图像处理中非常有用,而且还可以为在彩色图像处理这一领域继续研究提供一个必要的工具。伪彩色和以单独图像为基础的全彩色处理,提供了一个技术(这些技术在第3章到第5章较详细地阐述)纽带。

关于彩色向量空间的材料从以前已讨论过的方法出发,并突出灰度和全彩色处理间的重要区别。就技术而论,直接彩色向量处理这一领域范围很广并包括许多处理,如中值和其他类型滤波器、自适应和形态学滤波器、图像灰度复原、图像压缩和许多其他技术。这些处理对彩色图像分量进行的操作并不等效。下面列出的参考文献对该领域的进一步研究成果做了一个

介绍。

在彩色图像中，关于噪声的描述还指出向量问题的本质，遵循彩色图像例行地被从一个空间变换到另一空间的事实，问题的本质是如何减小图像中的噪声。在某些情况下，噪声的滤除可以在分量图像基础上进行，但另一方面，如中值滤波，需要就前边章节提到的彩色像素是向量这一点进行特殊处理。

虽然分割是第 10 章的主题，图像数据压缩是第 8 章的主题，但在彩色图像处理一章介绍它们有利于保持知识的连续性。如在随后的讨论中将明显看到的，后续各章所讨论的许多技术对本章研究的问题均可用。

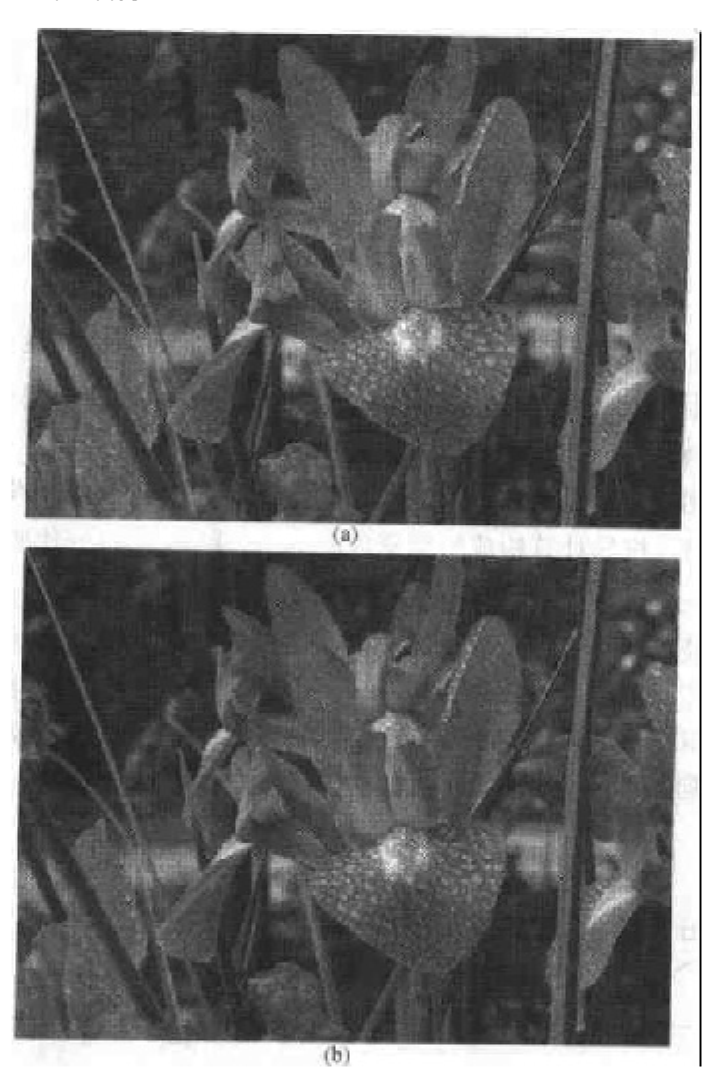


图 6.51 彩色图像压缩。(a)原 RGB 图像，(b)图像(a)的压缩解压缩结果

参考资料

对于彩色科学的全面参考，见Malacara[2001]。彩色生理学部分可见Gegenfurtne和Sharpe[1999]这两份参考文献，以及早期的Walsh[1958]和Kiver[1965]的书籍，该书就 6.1 节讨论的材料提供了丰富的补充材料。对彩色模型(6.2 节)的进一步读物，见Fortner和Meyer[1997]，

Poynton[1996]和Fairchild[1998]的文献。对6.2.3节中HSI模型的详细推导,见Smith[1978]的论文或咨询相关网站。伪彩色的内容(6.3节)紧密联系了数据可视化的一般领域。Wolff和Yaeger[1993]在伪彩色应用方面是很好的基础参考文献,Thorell和Smith[1990]的书也很重要。对于彩色信号向量表示方面的讨论(6.4节)见Plataniotis和Venetsanopoulos[2000]。

6.5节的参考文献有Benson[1985],Robertson[1977]和CIE[1978],还可参见MacAdam[1942]的经典论文。彩色图像滤波方面的材料(6.6节)是6.4节介绍的基于向量的公式及在第3章讨论的空间滤波。彩色图像分割(6.7节)是在过去10年备受关注的课题。Liu,Yang[1994]和Shafarenko等[1998]的论文是在这一领域有代表性的文献。IEEE图像处理学报的专刊[1997]也是重要资料。彩色边缘检测的讨论(6.7.3节)参见Di Zenzo[1986]。Plataniotis和Venetsanopoulos[2000]的书是对彩色图像分割各种方法的一个综述。6.8节的讨论以5.2节介绍的噪声模型为基础。图像压缩方面的文献(见6.9节)列在第8章的末尾。

习题

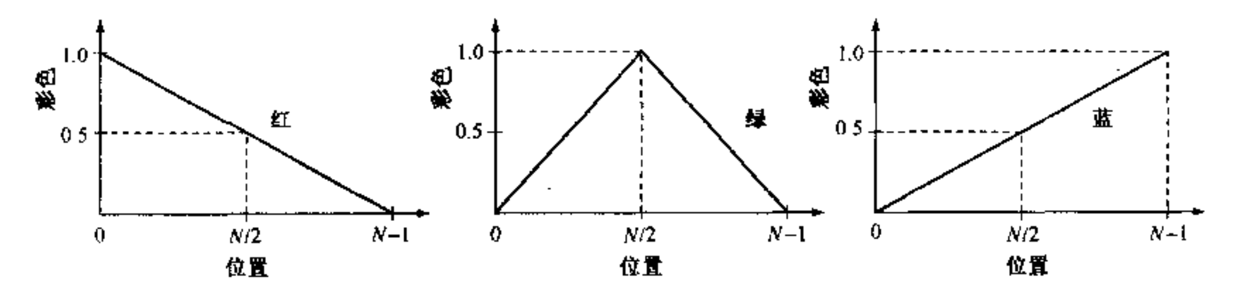
6.1 给出在图6.5中构成标为"暖白色"的点的红,绿和蓝的百分比。

★**6.2** 考虑任何两种有效的颜色 c_1 和 c_2,其在图6.5中色度图上的坐标为(x_1, y_1)和(x_2, y_2),推导出计算构成给定颜色的 c_1 和 c_2 百分比的必要的通用表达式,其中这种颜色位于连接这两种彩色的直线上。

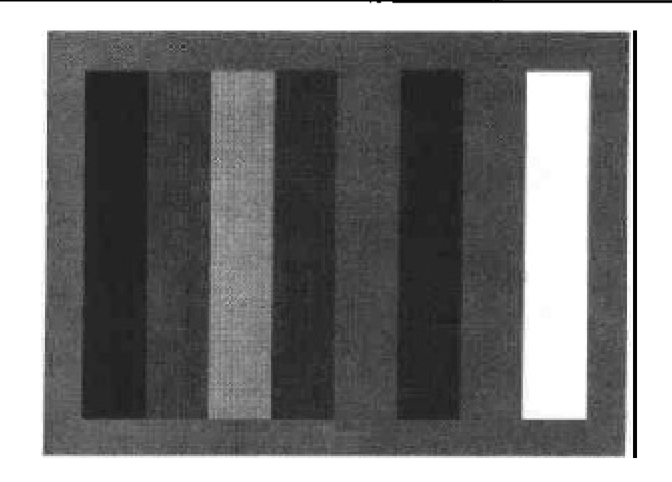
6.3 考虑任何可用的三种彩色 c_1, c_2 和 c_3,其在图6.5的色度图中坐标为(x_1, y_1),(x_2, y_2)和(x_3, y_3)。推导计算构成给定彩色的 c_1, c_2 和 c_3 的百分比所必需的通用表达式。该颜色位于三角形内,三角形的顶点在 c_1, c_2 和 c_3 的坐标处。

6.4 在一条自动装配线中,为了简化检测,三类零件被编以彩色码。然而,只有一台黑白电视摄像机可用于获取数字图像,提出一种用该摄像机检测3种不同颜色的技术。

6.5 在简单的RGB图像中,R, G 和 B 分量图像有示于下图的水平强度分布曲线。在该图像的中间一列人们能看到什么颜色?



6.6 当它们出现在单色监视器上时,做下列图像的RGB分量图。所有颜色都为最大强度与饱和度。在做这一问题时考虑作为图像一部分的中间灰度边界。



6.7 在彩色 RGB 系统中,每一 RGB 图像是一幅 8 比特图像,共有多少不同的灰度?

6.8 考虑示于图 6.8 的 RGB 彩色立方体并回答下列问题:

★(a) 描述一下在 R、G 和 B 原图像中灰度级怎样变化,原图像由彩色立方体的前表面组成。

(b) 假设用 CMY 彩色代替 RGB 立方体中的每一种颜色,这个新的立方体显示在 RGB监视器上,用一个彩色名标注在屏幕上看到的新立方体的 8 个顶点。

(c) 怎样描述涉及饱和度的 RGB 彩色立方体边缘上的彩色?

6.9 (a) 概略描述习题 6.6 中,图像出现在单色监视器上时 CMY 的分量。

(b) 如果在(a)中描述的 CMY 分量分别送入监视器的红,绿和蓝输入端,描述一下结果图像。

★**6.10** 从式(6.5.5)中的 RGB 对应物推导式(6.5.6)的 CMY 强度映射函数。

6.11 考虑示于图 6.10(a)的 216 安全色阵列,用其(行,列)号标注每一元素,左上角的元素是(1,1),最右边底部的元素为(12,18),在哪一点找到下列元素?

(a) 最纯的绿色。

(b) 最纯的蓝色。

★**6.12** 当它们出现在单色监视器上时,概略描述出习题 6.6 中图像的HSI分量。

6.13 为产生一个与示于缩放部分(在图 6.2 中叫做可见谱)的相似的彩色波段提供一个方法,注意该波段从左边的暗紫色开始并朝着右边的纯红色延伸(提示:用HSI彩色模型)。

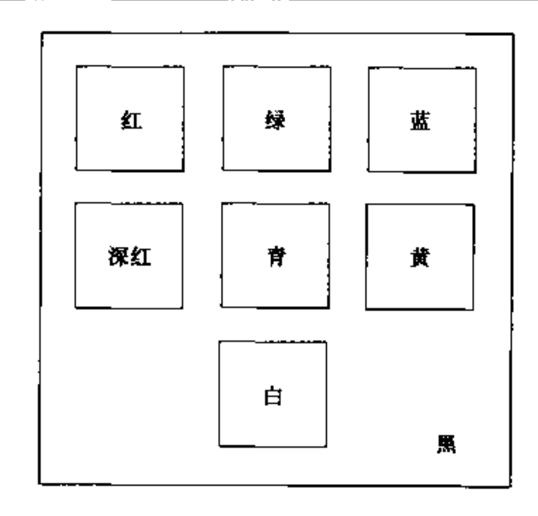
★**6.14** 为产生一种示于图 6.13(c)的图像的彩色方案,以流程图的方式给出你的答案,假定强度值固定且已给定(提示:用HSI彩色模型)。

6.15 考虑下面由实心方块组成的图像,为讨论你的答案,选择 8 种灰度组成灰度级,从 0 到 7,这里 0 表示黑,7 表示白。假定把图像转换到 HSI 彩色空间,如果它们有意义,对于色调等级用一个特殊的数字回答下列问题。否则用关系"与…相同"、"比…明亮"或"比…暗"等来满足上述要求,如果不能对你讨论的图像指定特殊的灰度级或这些关系之一,请给出原因。

(a) 概略描述色调图像。

(b) 概略叙述饱和度图像。

(c) 概略叙述强度图像。

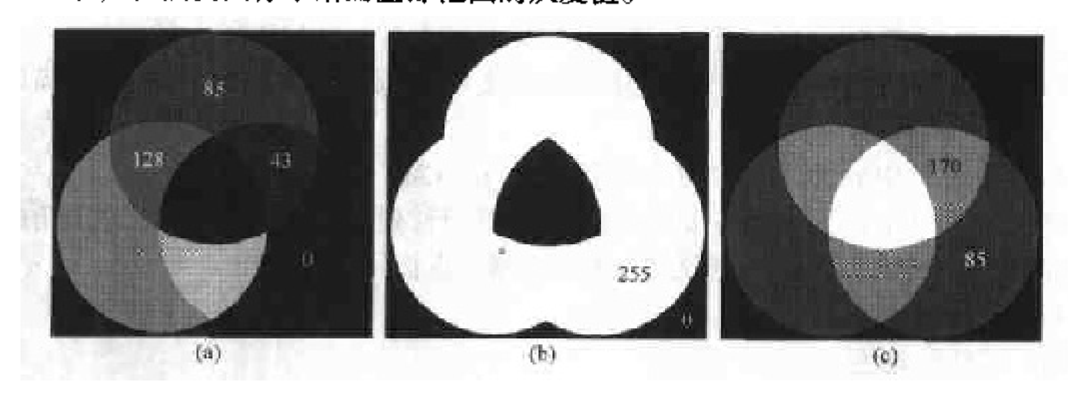


6.16 下列8比特图像(从左到右)是来自图6.16的 H,S,I 分量图像。数字指出灰度值。回答下列问题,解释你回答的根据。如果基于给定的信息不可能回答这一问题,叙述为什么不能回答。

★(a) 在色调图像中给出全部范围的灰度值。

(b) 在饱和度图像中给出全部范围的灰度值。

(c) 在强度图像中给出全部范围的灰度值。



6.17 参考图6.27回答下列问题:

(a) 为什么图6.27(f)中的图像主要呈现红色调。

(b) 提出一种自动编码程序,图6.27中水是亮蓝色。

(c) 提出一种自动编码程序,人工分量以亮黄色为主。

★6.18 说明彩色图像的补色饱和度分量不能单独从输入图像的饱和度分量计算出来。

6.19 说明使用HSI彩色模型对于图6.33(b)中的近似补色的色调变换函数形状。

★6.20 推导产生一幅彩色图像补色的CMY变换。

6.21 画出用以在RGB彩色空间校正对比度的变换函数形状。

★6.22 假定一个图像系统的监视器和打印机显示偏黄色,描述一下用来校正不平衡的通用变换。

6.23 在习题6.6假定的问题中,计算 $L^*a^*b^*$ 图像分量。

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.588 & 0.179 & 0.183 \\ 0.29 & 0.606 & 0.105 \\ 0 & 0.068 & 1.021 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

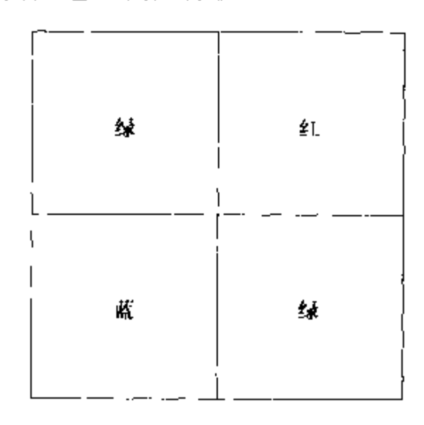
这个矩阵方程定义了由D65标准照明条件下(Benson[1985])观看的 NISC 美国彩色电视荧光粉产生的彩色三激励值。

★6.24 根据 3.3.2 节你怎样实现灰度级直方图匹配(规定)的彩色等效。

6.25 考虑下列 500×500 RGB 彩色图像,方框是红,绿,蓝。

(**a**) 假定将该图像转换到 HSI,用 25×25 平均模板模糊 H 分量图像,并转换回到 RGB,会看到什么结果?

(**b**) 重复上题,这次是模糊饱和度图像。



6.26 当 $\mathbf{C}=\mathbf{I}$(单位矩阵)时,说明式(6.7.2)简化为式(6.7.1),给出矩阵式。

6.27 ★(**a**)参考 6.7.2 节的讨论,给出一个过程(以流程图的形式)确定一个彩色向量点 $\mathbf{z}$ 是否在一个立方体内部,立方体宽度为 W,中心点在平均彩色向量 $\mathbf{a}$ 处,不允许计算距离。

(**b**)这个处理还可在图像对图像的基础上执行,如果盒子位于轴的上部,示出你如何去做。

6.28 描述一下在 RGB 彩色空间满足方程:

$$D(\mathbf{z},\mathbf{a}) = [(\mathbf{z}-\mathbf{a})^T\mathbf{C}^{-1}(\mathbf{z}-\mathbf{a})]^{\frac{1}{2}} = D_0$$

的那些点组成的表面。这里 D_0 是特定的非零常数,假定 $\mathbf{a}=\mathbf{0}$ 并且:

$$\mathbf{c} = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

6.29 参考 6.7.3 节,也许会想到,在任何点定义 RGB 图像梯度的逻辑方法是,计算每个分量图像的梯度向量(见 3.7.3 节),然后把 3 个单独的梯度向量求和形成彩色图像梯度向量。遗憾的是,有时这一方法会导致错误结果。特别是对明确定义了边缘并有 0 梯度的彩色图像,如果用这一方法容易产生错误。给出一幅这样图像的例子(提示:置一个彩色平面为恒定值,简化你的分析)。

第7章　小波变换和多分辨率处理

虽然从20世纪50年代末起傅里叶变换就一直是变换域图像处理的基石，近来，一种新的变换，称做小波变换，使得压缩、传输和分析许多图像变得更为便捷。不像傅里叶变换，它的基础函数是正弦函数，与此不同，小波变换基于一些小型波，称为小波，具有变化的频率和有限的持续时间。这就允许它们对图像提供一张等效的乐谱，不光阐明了要演奏的音符(或频率)，而且阐明了何时要演奏。而另一方面，传统的傅里叶变换，只提供了音符或频率信息，局部信息在变换过程中丢失了。

1987年，在一种全新而有效的信号处理与分析方法——多分辨率理论(Mallat[1987])中，小波首次作为分析基础出现了。多分辨率理论将多种学科的技术有效地统一在一起，如信号处理的子带编码、数字语音识别的积分镜像过滤以及金字塔图像处理。就像它的名字所表达的那样，多分辨率理论与多种分辨率下的信号(或图像)表示和分析有关。其优势很明显——某种分辨率下所无法发现的特性在另一种分辨率下将很容易被发现。尽管20世纪80年代末以前图像处理界并没有对多分辨率分析投注太多的热情，现在关于该专题的论文和专著却是多得数不胜数。

本章中，将从多分辨率的角度来审视小波变换。虽然解释小波变换的方式有很多，但这种方式能简化数学和物理的解释过程。将从影响多分辨率理论形成的图像技术概述入手，目的在于在图像处理的环境中阐述该理论的基础概念，并同时对该方法及其应用做一个简要的历史回顾。本章的主要内容将集中在用于图像表示和分析方法的多分辨率工具箱的建立。为了更好地阐明该工具箱的用途，还给出了一些应用实例，如图像编码，噪声去除和边缘提取等。下一章中，小波将主要用于图像压缩，在此方向的应用已得到了广泛的重视。

7.1　背景

当观察图像时，通常看到的是相连接的纹理与灰度级相似的区域，它们相结合形成物体。如果物体的尺寸很小或对比度不高，通常采用较高的分辨率观察；如果物体尺寸很大或对比很强，只需要较低的分辨率。如果物体尺寸有大有小，或对比有强有弱的情况同时存在，以若干分辨率对它们进行研究将具有优势。当然，这就是多分辨率处理的魅力所在。

从数学的观点看，图像是一个亮度值的二维矩阵，像边界和对比强烈区域那样的突变特性的不同组合会产生统计值的局部变化。如图7.1所示(该图像将在本章的其他部分多次出现)，在同一图像的不同部分，即使是一阶统计值也会大不相同，因此，无法对整个图像定义一个简单的统计模型。

7.1.1　图像金字塔

以多分辨率来解释图像的一种有效但概念简单的结构就是图像金字塔(Burt和Adelson[1983])。图像金字塔最初用于机器视觉和图像压缩，一幅图像的金字塔是一系列以金字塔形

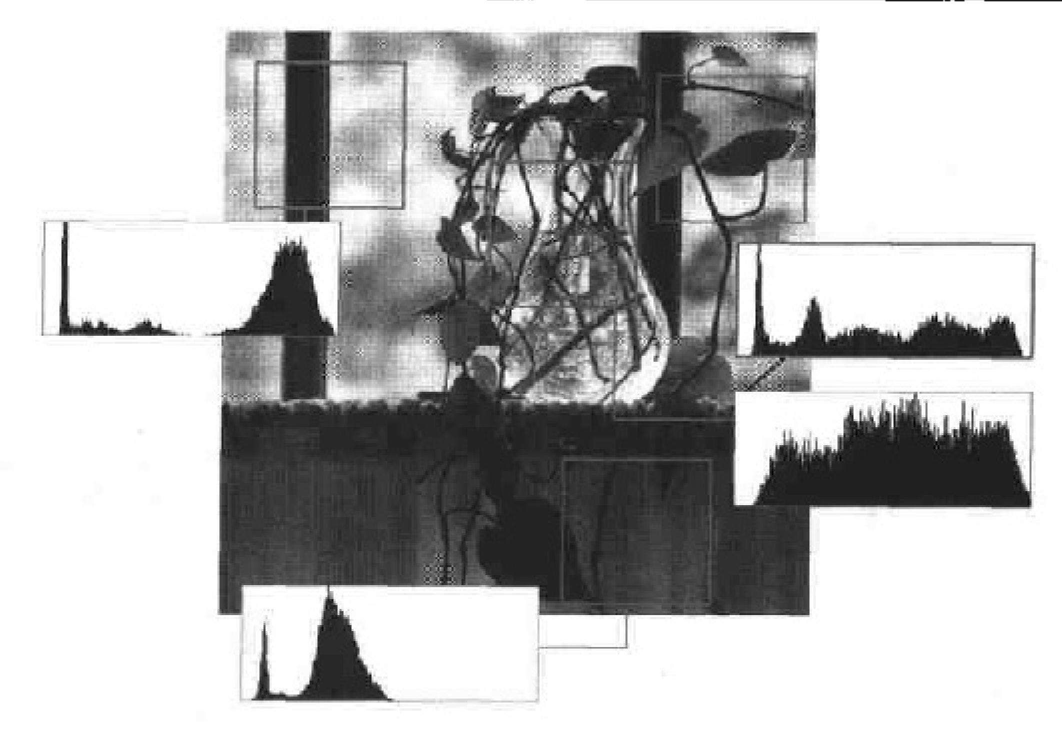


图 7.1　一幅自然图像和它的局部直方图变化

状排列的分辨率逐步降低的图像集合。如图 7.2(a)所示，金字塔的底部是待处理图像的高分辨率表示，而顶部是低分辨率的近似。当向金字塔的上层移动时，尺寸和分辨率就降低。因为基础级 J 的尺寸是 $2^J \times 2^J$ 或 $N \times N(J=\log_2 N)$，中间级 j 的尺寸是 $2^j \times 2^j$，其中 $0 \leqslant j \leqslant J$。完整的金字塔由 $J+1$ 个分辨率级组成，由 $2^J \times 2^J$ 到 $2^0 \times 2^0$，但大部分金字塔只有 $P+1$ 级，其中 $j=J-P,\cdots,J-2,J-1,J$ 且 $1 \leqslant P \leqslant J$。也就是说，通常限制它们只使用 P 级来减少原始图像近似值的尺寸。例如，一幅 512×512 图像的 1×1 或单像素近似值将非常小。$P+1$级金字塔($P>0$)中的元素总数是：

$$N^2\left(1+\frac{1}{(4)^1}+\frac{1}{(4)^2}+\cdots+\frac{1}{(4)^P}\right) \leqslant \frac{4}{3}N^2$$

图 7.2(b)显示了一个建立图像金字塔的简单系统。$j-1$ 级的近似输出用来建立近似值金字塔，包括原始图像的一个或多个近似值。作为金字塔基级的原始图像和它的 P 级减少的分辨率近似都能直接获取并调整。图 7.2(b)中 j 级的预测残差输出用于建立预测残差金字塔。这些金字塔包括了原始图像的 $J-P$ 级低分辨率的近似信息，以及建立 P 级较高分辨率的近似信息。j 级的信息在相应近似金字塔的 j 级近似与基于 $j-1$ 级预测残差得到的近似估计之间是不同的。对这些差异进行编码(用于存储或传输)将比对近似值进行编码有效得多。

如图 7.2(b)的框图所表明的，近似值和预测残差金字塔都是以一种迭代的方式进行计算的。$P+1$ 级金字塔通过执行 P 次框图中的操作建立。第一次迭代和传递时，$j=J$，并且 $2^J \times 2^J$ 的原始图像作为 J 级的输入图像，从而产生 $J-1$ 级近似值和 J 级预测残差。对于$j=J-1, J-2,\cdots,J-P+1$(按这一顺序)的传递，前面迭代的 $j-1$ 级近似值输出将作为输入。每次传递由 3 个连续步骤组成：

1. 计算输入图像的减少的分辨率近似值。这可以通过对输入进行滤波并以 2 为步长进行抽样去做(即子抽样)。可以采用的滤波操作有很多,如邻域平均(它可生成平均值金字塔),高斯低通滤波器(它可生成高斯金字塔),或者不进行滤波,生成子抽样金字塔。生成近似值的质量是所选滤波器的函数,在图 7.2(b)中以 $j-1$ 级近似值进行标记。没有滤波器,在金字塔的上一层混淆变得很显著,子抽样点对所采取的区域没有很好的代表性。
2. 对上一步的输出进行内插——因子仍为 2——并进行过滤。这将生成与输入等分辨率的预测图像。由于在步骤 1 的输出像素之间进行插值运算,插入滤波器决定了预测值与步骤 1 的输入之间的近似程度。如果插入滤波器被忽略了,预测值将是步骤 1 输出的内插形式,复制像素的块效应将变得很明显。
3. 计算步骤 2 的预测值和步骤 1 的输入之间的差异。以 j 级预测残差进行标识的这个差异将用于原始图像的重建(见例 7.1)。在没有量化差异的情况下,预测残差金字塔可以用于生成相应的近似金字塔,包括原始图像,而没有误差。

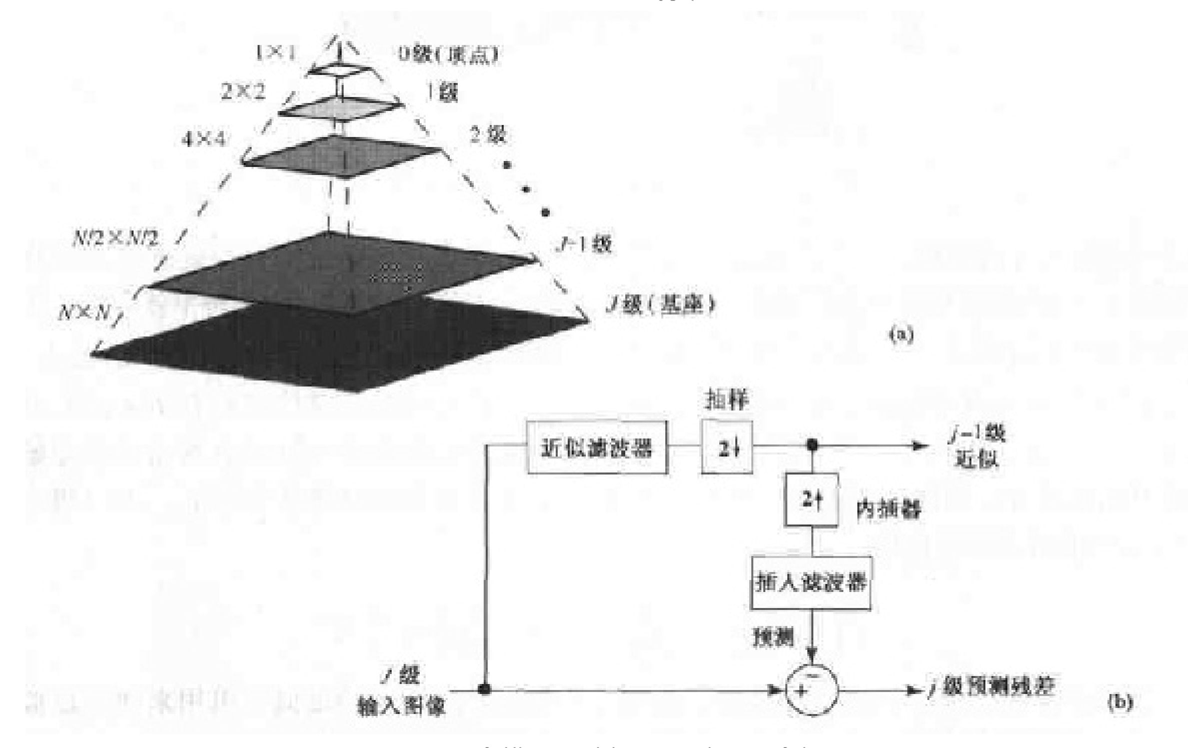


图 7.2　(a)一个金字塔图像结构,(b) 建立金字塔的方框图

执行上述过程 P 次将产生密切相关的 $P+1$ 级近似值和预测残差金字塔。$j-1$ 级近似值的输出用于提供近似值金字塔;而 j 级预测残差的输出被放在预测残差金字塔中。如果不需要预测残差金字塔,步骤 2、3 和内插器、插入滤波器以及图 7.2(b)中的加法器都可以省略。

例 7.1　高斯和拉普拉斯金字塔

图 7.3 显示了图 7.1 中花瓶的一种可能的近似值和预测残差金字塔。图 7.3(a)中的近似值金字塔是一个高斯金字塔。使用 4.2.4 节的图 4.9(c)所描述的 5×5 低通高斯卷积核在空间域进行过滤。正如所看到的那样,处理后的金字塔包括 512×512 分辨率的原始图像(在底部)和 3 个低分辨率的近似值(分辨率分别是 256×256, 128×128 以及 64×64)。即 $P=3$ 且金字塔被缩减到 4 级——可能的 $\log_2(512)+1$ 级(或者 10 级)中的 9,8,7,6 级。

注意,金字塔的分辨率越低,伴随的细节越少。例如,第6级(即64×64)近似对于定位窗框很合适,但对于寻找黄蘖的茎就很不合适了。通常金字塔的低分辨率图像用于分析大的结构或图像的整体内容,而高分辨率图像用于分析单个物体的特性。这样的由粗糙到精细的分析策略在模式识别中特别适用。

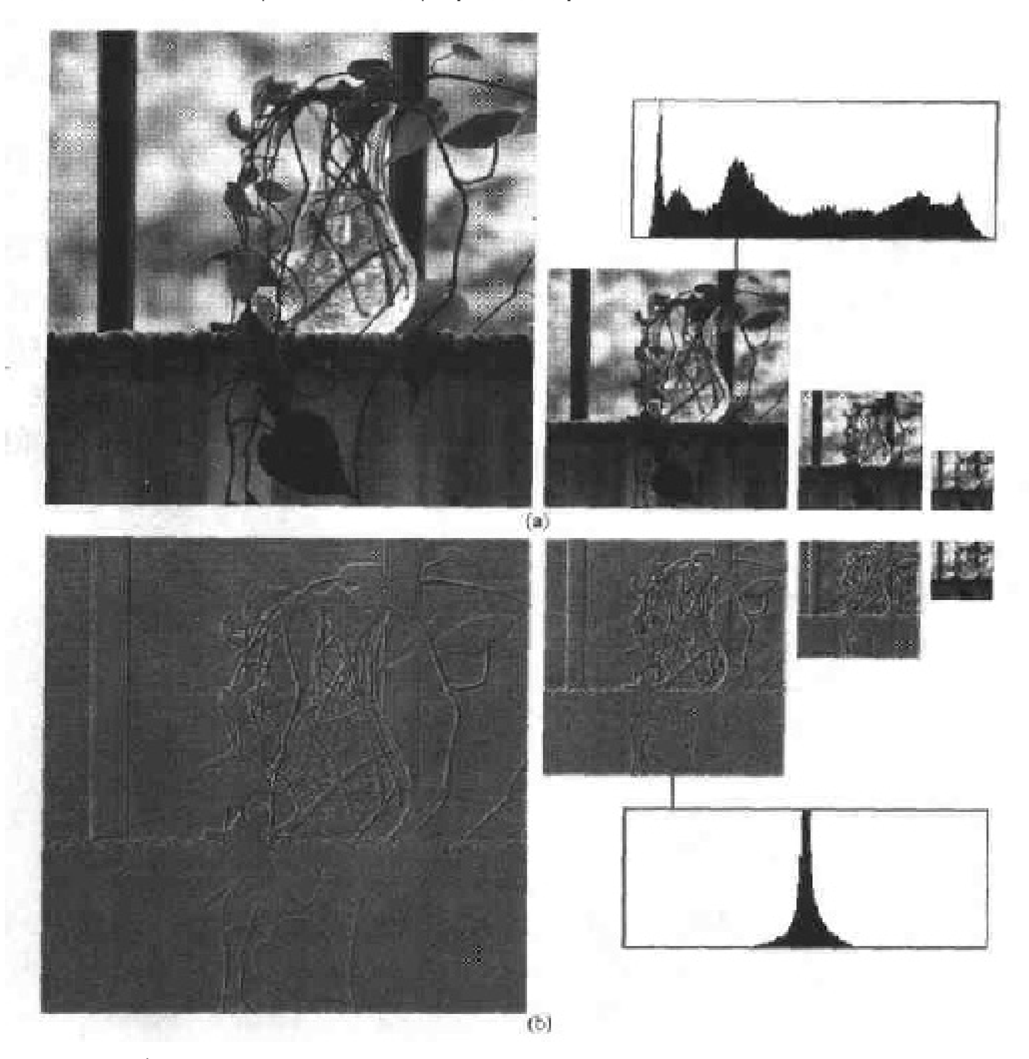


图7.3　两种图像金字塔和它的统计特性。(a)高斯金字塔(近似),(b)拉普拉斯金字塔(预测残差)

图7.3(b)中的拉普拉斯金字塔包含了用于计算7.3(a)中其高斯对应部分的预测残差。为建立高斯金字塔,首先从拉普拉斯金字塔的第6级,64×64近似图像开始预测高斯金字塔的第7级,分辨率128×128的近似值(通过内插和滤波实现),并加上拉普拉斯的第7级预测残差。这个过程被重复使用计算近似图像,直到生成512×512的原始图像。注意,拉普拉斯金字塔的预测残差图像的一阶统计值是零点附近的高峰值。与它们的高斯对应部分不同,这些图像可以通过为更多可能值分配较少比特数实现高比例压缩(参见8.1.1节的变长编码)。最终,将注意到图7.3(b)中的预测残差达到一种使较小的预测误差更明显的程度。然而,预测误差的直方图却是基于预标定的残差,以第128级表示0误差。

7.1.2 子带编码

另一种与多分辨率分析相关的重要图像技术是子带编码。在子带编码中,一幅图像被分解成为一系列限带分量的集合,称为子带,它们可以重组在一起无失真地重建原始图像。最初是为语音和图像压缩而研制的,每个子带通过对输入进行带通滤波而得到。因为所得到的子带带宽要比原始图像的带宽小,子带可以进行无信息损失的抽样。原始图像的重建可以通过内插、滤波和叠加单个子带来完成。

图7.4(a)显示了两段子带编译码系统的基本部分。系统的输入是一个一维的带限时间离散信号 $x(n), n=0,1,2,\cdots$;输出序列 $\hat{x}(n)$ 是通过分析滤波器 $h_0(n)$ 和 $h_1(n)$ 将 $x(n)$ 分解成 $y_0(n)$ 和 $y_1(n)$,然后再通过综合滤波器 $g_0(n)$ 和 $g_1(n)$ 综合得到的。注意,滤波器 $h_0(n)$ 和 $h_1(n)$ 是半波数字滤波器,其理想传递函数 H_0 和 H_1 如图7.4(b)所示。滤波器 H_0 是一个低通滤波器,输出是 $x(n)$ 的近似值;滤波器 H_1 是一个高通滤波器,输出是 $x(n)$ 的高频或细节部分。所有的滤波都通过在时间域将每个滤波器的输入与其冲激响应(对单位强度冲激函数 $\delta(n)$ 的响应)进行卷积来实现。我们希望能通过选择 $h_0(n), h_1(n), g_0(n)$ 和 $g_1(n)$(或 H_0, H_1, G_0 和 G_1)来实现对输入的完美重构,即 $\hat{x}(n)=x(n)$。

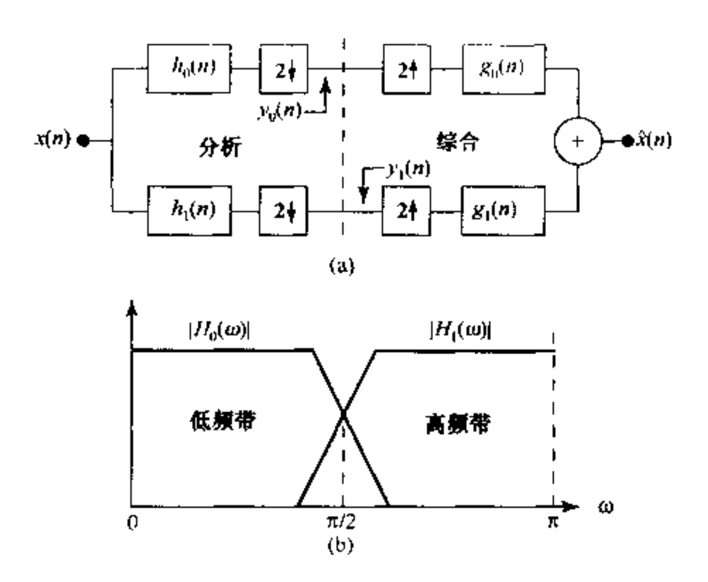


图7.4　(a)一维子带编码和解码的两频带滤波器组,(b)频谱分离特性

离散傅里叶变换的一般推广(Z 变换)是研究如图7.4(a)所示的离散时间、数据采样系统的理想工具。序列 $x(n)(n=0,1,2,\cdots)$ 的 Z 变换是:

$$X(z)=\sum_{-\infty}^{\infty}x(n)z^{-n}\tag{7.1.1}$$

其中 z 是一个复变量[如果用 $e^{j\omega}$ 代替 z,式(7.1.1)将成为 $x(n)$ 的离散傅里叶变换]。对 Z 变换的兴趣源于它处理采样率变化的便捷性。时域中以2为因子的抽样对应到 Z 域中为:

$$x_{\text{down}}(n)=x(2n)\Leftrightarrow X_{\text{down}}(z)=\frac{1}{2}[X(z^{1/2})+X(-z^{1/2})]\tag{7.1.2}$$

其中,双箭头表示左右两端的表达式组成了一对 Z 变换对。同样,以2为因子的内插,可以由变换对定义为:

$$x^{up}(n) = \begin{cases} x(n/2) & n = 0,2,4,\cdots \\ 0 & 其他 \end{cases} \Leftrightarrow X^{up}(z) = X(z^2) \quad (7.1.3)$$

如果对 $x(n)$先抽样再内插得到 $\hat{x}(n)$,由式(7.1.2)和式(7.1.3)结合可得:

$$\hat{X}(z) = \frac{1}{2}[X(z) + X(-z)] \quad (7.1.4)$$

其中 $\hat{x}(n) = Z^{-1}[\hat{X}(z)]$就是对序列抽样-内插得到的结果。这个等式中$X(-z)$项是序列 $x(n)$混叠的或调制的 Z 变换。其反 Z 变换是:

$$Z^{-1}[X(-z)] = (-1)^n x(n) \quad (7.1.5)$$

根据上述 Z 变换简介,再来考虑图 7.4(a)中的子带编译码系统。由式(7.1.4)可将系统输出表达为:

$$\hat{X}(z) = \frac{1}{2}G_0(z)[H_0(z)X(z) + H_0(-z)X(-z)] \quad (7.1.6)$$

$$+ \frac{1}{2}G_1(z)[H_1(z)X(z) + H_1(-z)X(-z)]$$

其中,图 7.4(a)中的滤波器 $h_0(n)$的输出由下述变换对定义:

$$h_0(n) * x(n) = \sum_k h_0(n-k)x(k) \Leftrightarrow H_0(z)X(z) \quad (7.1.7)$$

与傅里叶变换一样,时域(或空间域)的卷积等价于 Z 域的乘积。整理式(7.1.6)可得:

$$\hat{X}(z) = \frac{1}{2}[H_0(z)G_0(z) + H_1(z)G_1(z)]X(z) \quad (7.1.8)$$

$$+ \frac{1}{2}[H_0(-z)G_0(z) + H_1(-z)G_1(z)]X(-z)$$

其中,第二项由于含有 $-z$ 的关系,它代表了抽样-内插过程带来的混叠。

对于输入的无失真重建,$\hat{x}(n) = x(n)$和 $\hat{X}(z) = X(z)$。因此,可以假定下列条件:

$$H_0(-z)G_0(z) + H_1(-z)G_1(z) = 0 \quad (7.1.9)$$

$$H_0(z)G_0(z) + H_1(z)G_1(z) = 2 \quad (7.1.10)$$

式(7.1.9)通过强制式(7.1.8)的第二项为零消除了混叠;而式(7.1.10)通过强制第一项等于 $X(z)$消除了幅度失真。二者都可以合并成一维矩阵表达式:

$$[G_0(z) \quad G_1(z)]\mathbf{H}_m(z) = [2 \quad 0] \quad (7.1.11)$$

其中分析调制矩阵 $\mathbf{H}_m(z)$为:

$$\mathbf{H}_m(z) = \begin{bmatrix} H_0(z) & H_0(-z) \\ H_1(z) & H_1(-z) \end{bmatrix} \quad (7.1.12)$$

假定 $\mathbf{H}_m(z)$是非奇异矩阵(即左右倒置),整理式(7.1.11),左乘$(\mathbf{H}_m^T(z))^{-1}$可得:

$$\begin{bmatrix} G_0(z) \\ G_1(z) \end{bmatrix} = \frac{2}{\det(\mathbf{H}_m(z))}\begin{bmatrix} H_1(-z) \\ -H_0(-z) \end{bmatrix} \quad (7.1.13)$$

$\det(\mathbf{H}_m(z))$表示 $\mathbf{H}_m(z)$的行列式。

式(7.1.9)到式(7.1.13)揭示了重建滤波器组的若干重要特性。如,矩阵(7.1.13)告诉我们 $G_1(z)$是 $H_0(-z)$的函数,而 $G_0(z)$是 $H_1(-z)$的函数。分析和综合滤波器交叉调制,也就是说,在 7.4(a)的框图中对角线上相对的滤波器在 Z 域中以 $-z$ 相关联。对于有限冲激响应(FIR)滤波器,调制矩阵的行列式是一个纯时延,即 $\det(\mathbf{H}_m(z)) = \alpha z^{-(2k+1)}$(见 Vetterli 和 Kovacevic[1995])。因此,交叉调制的准确形式是 α 的函数。$z^{-(2k+1)}$项可被认为是任意的,因为它只改变滤波器的群时延。忽略时延,令 $\alpha = 2$,对式(7.1.13)做反 Z 变换,可得到:

$$g_0(n) = (-1)^n h_1(n) \tag{7.1.14}$$
$$g_1(n) = (-1)^{n+1} h_0(n)$$

如果 $\alpha = -2$,结果的表达式符号相反:

$$g_0(n) = (-1)^{n+1} h_1(n) \tag{7.1.15}$$
$$g_1(n) = (-1)^n h_0(n)$$

因此,FIR 综合滤波器是分析滤波器的交叉调制的副本,有且仅有一个符号相反。

式(7.1.9)到式(7.1.13)也可以用来证明分析和综合滤波器的双正交性。令低通分析滤波器和综合滤波器传递函数的乘积为 $P(z)$。由式(7.1.13)代入 G_0,可得:

$$P(z) = G_0(z)H_0(z) = \frac{2}{\det(\mathbf{H}_m(z))}H_0(z)H_1(-z) \tag{7.1.16}$$

由于 $\det(\mathbf{H}_m(z)) = -\det(\mathbf{H}_m(-z))$,

$$G_1(z)H_1(z) = \frac{-2}{\det(\mathbf{H}_m(z))}H_0(-z)H_1(z) = P(-z) \tag{7.1.17}$$

因此,$G_1(z)H_1(z) = P(-z) = G_0(-z)H_0(-z)$,式(7.1.10)变成:

$$G_0(z)H_0(z) + G_0(-z)H_0(-z) = 2 \tag{7.1.18}$$

做反 Z 变换可得:

$$\sum_k g_0(k)h_0(n-k) + (-1)^n \sum_k g_0(k)h_0(n-k) = 2\delta(n)$$

通常,冲激函数 $\delta(n)$在 $n=0$ 时等于 1,而在其他情况下等于 0。由于奇次方项相互抵消,可得[①]:

$$\sum_k g_0(k)h_0(2n-k) = \langle g_0(k), h_0(2n-k)\rangle = \delta(n) \tag{7.1.19}$$

由式(7.1.9)和式(7.1.10)开始,并将 H_0 和 G_0 表示成 G_1 和 H_1 的函数,可得:

$$\langle g_1(k), h_1(2n-k)\rangle = \delta(n) \tag{7.1.20}$$
$$\langle g_0(k), h_1(2n-k)\rangle = 0$$

且

$$\langle g_1(k), h_0(2n-k)\rangle = 0$$

① 序列 $x(n)$ 和 $y(n)$ 的向量内积是 $\langle x, g\rangle = \sum_n x^*(n)y(n)$,这里 * 号表示复共轭操作。如果 $x(n)$ 和 $y(n)$ 是实数,$\langle x, y\rangle = \langle y, x\rangle$。

与式(7.1.19)合并,可得到更有普遍意义的表达式:

$$\langle h_i(2n-k), g_j(k)\rangle = \delta(i-j)\delta(n), i,j = \{0,1\} \tag{7.1.21}$$

满足该条件的滤波器组称为具有双正交性。此外,所有两频段实系数的完美重建滤波器组的分析和综合滤波器的冲激响应服从双正交性约束。双正交 FIR 滤波器的例子有双正交spline族(Cohen,Daubechies和Feauveau[1992])和双正交coiflet族(Tian和Wells[1995])。

表 7.1 给出了式(7.1.9)和式(7.1.10)的通解。虽然它们都能满足式(7.1.21)的双正交要求,但各自的求解方式不同,定义的可完美重建的滤波器类也不同。每类中都依一定规格设计了一个"原型"滤波器,而其他滤波器由原型计算产生。表 7.1 的 1,2 列是滤波器族文献的经典结果,称为正交镜像滤波器(QMF)(Croisier,Estaban和Galand[1976])和共轭正交滤波器(CQF)(Smith和Barnwell[1986])。第 3 列中的滤波器称为具有正交性,用于后面快速小波变换的开发中(见 7.4 节)。它们在双正交的基础上更进一步,要求:

$$\langle g_i(n), g_j(n+2m)\rangle = \delta(i-j)\delta(m), i,j = \{0,1\} \tag{7.1.22}$$

这对可完美重建的滤波器族定义了正交性。注意,表中第 3 行 $G_1(z)$的表达式。$2K$ 代表各滤波器系数的长度或数目(即滤波器抽头)。可见,G_1 与低通综合滤波器 G_0 的联系在于调制[见式(7.1.5)]、时域反转或奇数平移①。此外,H_0 和 H_1分别是相应综合滤波器 G_0 和 G_1 的时域反转。从表 7.1 的第 3 列中选取适当的输入,做反 Z 变换,可得:

$$g_1(n) = (-1)^n g_0(2K-1-n) \tag{7.1.23}$$
$$h_i(n) = g_i(2K-1-n), i = \{0,1\}$$

其中,h_0,h_1,g_0 和 g_1 是定义的正交滤波器的冲激响应,如Smith和Barnwell滤波器(Smith和Barnwell[1984]),Daubechies滤波器(Daubechies[1988]),Vaidyanathan和Hoang滤波器(Vaidyanathan和Hoang[1988])。

表 7.1 中的一维滤波器也可用于图像处理的二维可分离滤波器。如图 7.5 所示,可分离滤波器首先应用于某一维(如垂直向),再应用于另一维(如水平向)。此外,抽样也分两步执行——在第二次滤波前执行一次以减少计算量。滤波后的输出结果,用图 7.5 中的 $a(m,n)$,$d^V(m,n)$,$d^H(m,n)$和 $d^D(m,n)$表示,分别称为近似值、垂直细节、水平细节和图像的对角线细节子带。一个或多个这样的子带可被分为 4 个更小的子带,并可重复划分。

表 7.1　完美重建滤波器族

滤波器	QMF	CQF	正　交
$H_0(z)$	$H_0^2(z) - H_0^2(-z) = 2$	$H_0(z)H_0(z^{-1}) + H_0^2(-z)H_0(-z^{-1}) = 2$	$G_0(z^{-1})$
$H_1(z)$	$H_0(-z)$	$z^{-1}H_0(-z^{-1})$	$G_1(z^{-1})$
$G_0(z)$	$H_0(z)$	$H_0(z^{-1})$	$G_0(z)G_0(z^{-1}) + G_0(-z)G_0(-z^{-1}) = 2$
$G_1(z)$	$-H_0(-z)$	$zH_0(-z)$	$-z^{-2K+1}G_0(-z^{-1})$

① 对时间反转和平移的 z 变换对分别是 $x(-n) \Leftrightarrow X(z^{-1})$和 $x(n-k) \Leftrightarrow z^{-k}X(z)$。

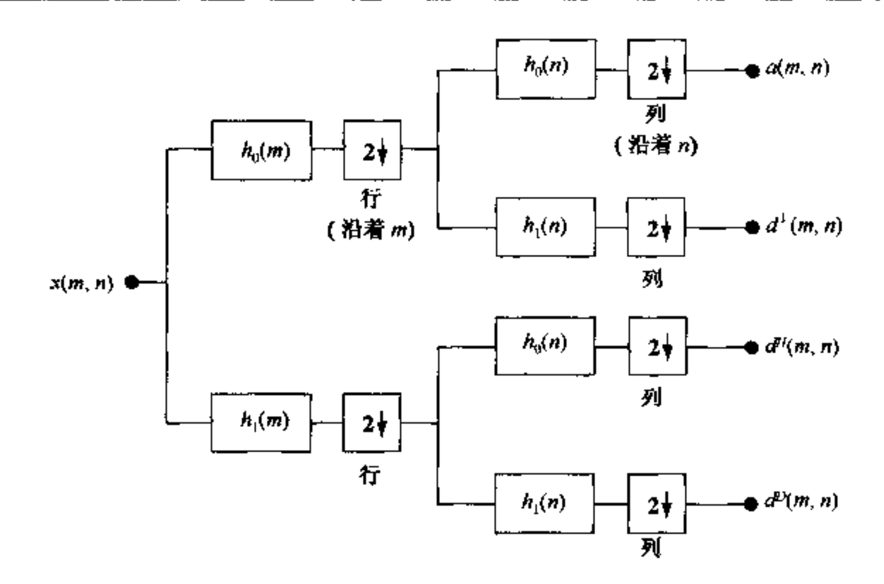


图 7.5 子带图像编码的二维 4 频段滤波器组

例 7.2 图 7.1 中花瓶的 4 频段子带编码

图 7.6 显示了一个 8 抽头正交滤波器的冲激响应。低通滤波器 $h_0(n)(0 \leqslant n \leqslant 7)$ 的系数是 −0.010 597 40, 0.032 883 01, 0.030 841 38, −0.187 034 81, −0.027 983 76, 0.630 880 76, 0.714 846 57和 0.230 377 81 (Daubechies [1992]); 其余正交滤波器的系数可以通过式(7.1.23)计算。注意,图 7.6 中分析和综合滤波器的交叉调制。用数字计算来说明这些滤波器既是双正交的[满足式(7.1.21)]又是正交的[满足式(7.1.22)]相对容易。此外,它们也满足式(7.1.9)和式(7.1.10),支持已分解的输入的无误差重建。

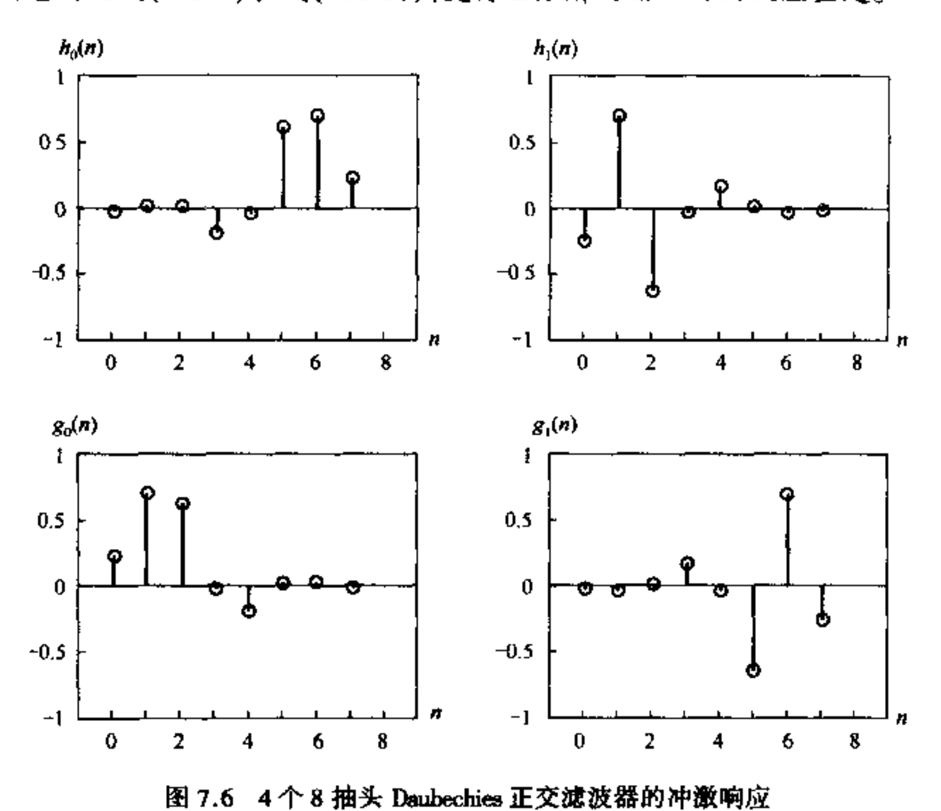


图 7.6 4 个 8 抽头 Daubechies 正交滤波器的冲激响应

图 7.7 显示了图 7.1 中花瓶的 512×512 图像基于图 7.6 中滤波器的 4 频段分离。该图像的每一个象限都是 256×256 的一个子带。从左上角开始，按顺时针方向旋转，4 个象限分别包括近似子带 a，水平细节子带 d^H，对角细节子带 d^D 和垂直细节子带 d^V。除了左上角的近似子带，所有子带都经过标定以使其基本结构更为明显。注意 d^H 和 d^V 中所表现出的混叠，这是由于对图 7.1 仅可分辨的窗口进行抽样造成的。根据式(7.1.9)，通过综合滤波器 $g_0(n)$和 $g_1(n)$实现的由子带完成的原始图像重建将消除这些混叠。为执行重建，需要图 7.5 系统的严格镜像滤波器组。在新的滤波器组中，$h_i(n)$，$i=\{0,1\}$，被它们的 $g_i(n)$对应部分所取代，并添加了内插器和加法器。

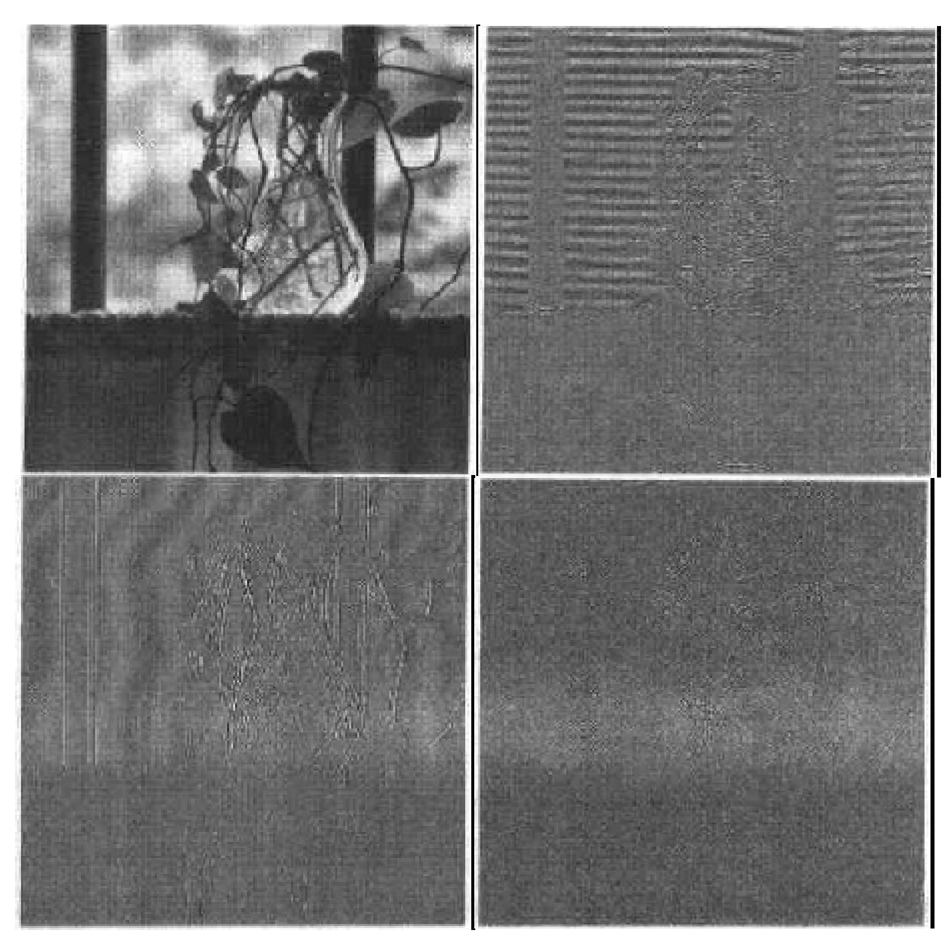


图 7.7　用图 7.5 的子带编码系统对图 7.1 中花瓶的 4 频段分离

7.1.3　哈尔变换

将要了解的第三个和最后一个与多分辨率分析有关的图像处理手段是哈尔变换(Haar [1910])。本章的内容中，其重要性体现在它的基函数是众所周知的最古老也是最简单的正交小波。下述部分将给出若干应用范例。

哈尔变换本身是可分离的，也是对称的，可以用下述矩阵形式表达：

$$\mathbf{T} = \mathbf{HFH} \tag{7.1.24}$$

其中，$\mathbf{F}$ 是一个 $N\times N$ 图像矩阵，$\mathbf{H}$ 是 $N\times N$ 变换矩阵，$\mathbf{T}$ 是 $N\times N$ 变换的结果。对于哈尔

变换,变换矩阵 **H** 包含哈尔基函数 $h_k(z)$,它们定义在连续闭区间 $z\in[0,1]$,$k=0,1,2,\cdots,N-1$,这里 $N=2^n$。为生成 **H** 矩阵,定义整数 k ,即 $k=2^p+q-1$(这里 $0\leqslant p\leqslant n-1$,$p=0$ 时,$q=0$ 或 1,$p\neq 0$ 时,$1\leqslant q\leqslant 2^p$)。可得哈尔基函数为:

$$h_0(z)=h_{00}(z)=\frac{1}{\sqrt{N}},z\in[0,1] \tag{7.1.25}$$

且

$$h_k(z)=h_{pq}(z)=\frac{1}{\sqrt{N}}\begin{cases}2^{p/2} & (q-1)/2^p\leqslant z<(q-0.5)/2^p\\ -2^{p/2} & (q-0.5)/2^p\leqslant z<q/2^p\\ 0 & 其他,z\in[0,1]\end{cases} \tag{7.1.26}$$

$N\times N$ 哈尔变换矩阵的第 i 行包含了元素 $h_i(z)$,其中 $z=0/N,1/N,2/N,\cdots,(N-1)/N$。例如,$N=4$ 时,k,q 和 p 值如下:

k	p	q
0	0	0
1	0	1
2	1	1
3	1	2

则 4×4 变换矩阵 $\mathbf{H}_4$ 为:

$$\mathbf{H}_4=\frac{1}{\sqrt{4}}\begin{bmatrix}1 & 1 & 1 & 1\\ 2 & 1 & -1 & -1\\ \sqrt{2} & -\sqrt{2} & 0 & 0\\ 0 & 0 & \sqrt{2} & -\sqrt{2}\end{bmatrix} \tag{7.1.27}$$

类似地,2×2 变换矩阵 $\mathbf{H}_2$ 是:

$$\mathbf{H}_2=\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} \tag{7.1.28}$$

它的基函数仅定义了 2 抽头 FIR 滤波器族,可满足表 7.1 中第一行第一列的 QMF 滤波器原型的规范。相应 QMF 分析滤波器 $h_0(n)$和 $h_1(n)$的系数分别是矩阵 $\mathbf{H}_2$ 的第一行和第二行的元素。

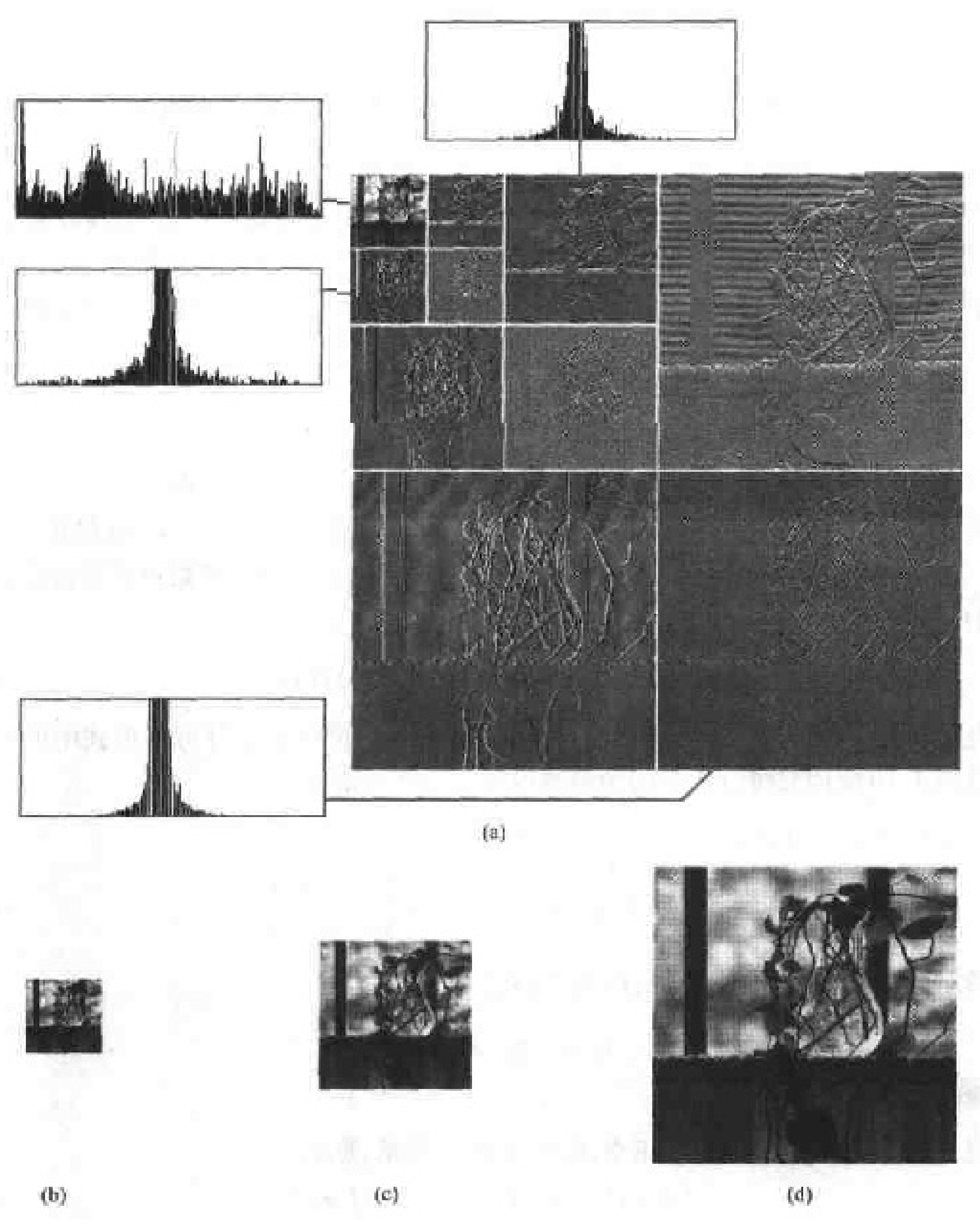
例 7.3　离散小波变换的哈尔函数

图 7.8(a)显示了使用式(7.1.25)和式(7.1.26)的哈尔基函数对图 7.1 的多分辨率分解。与图 7.3 中的金字塔形结构不同,该表示(被称为离散小波变换,将在本章的后续部分研究)包含了与原始图像同样的像素数。此外,

1. 其局部统计数据相对稳定并较易给出模型,见图 7.8(a)。
2. 其大多数值都接近 0,这使它对于图像压缩是一个优秀的候选者。
3. 原始图像的粗和细分辨率近似可以从中提取。图 7.8(b)到(d)即是从图 7.8(a)的子图像重建得来的。

在数据库的应用中，这些特性使得用户能够在搜索时很容易找到图像的低分辨率版本，然后再恢复附加数据以对其进行修饰。

最后，注意到，图 7.8(a)与图 7.7 的子带编码以及图 7.3(b)的拉普拉斯金字塔很相似。正如前面的结果，对图 7.8(a)的子图像进行标定以使它们的基本结构更明显。图 7.8(b)到(d)的近似图像尺寸分别是 64×64，128×128 和 256×256。对原始图像的 512×512 重建也完全可能。



(a)

(b)　(c)　(d)

图 7.8　(a)用哈尔基函数的离散小波变换，还显示了它的局部直方图变化，(b)～(d)一些不同的近似(64×64，128×128 和 256×256)可以从(a)得到

7.2 多分辨率展开

前述部分介绍了三种著名的图像处理技术,它们在数学理论多分辨率分析(MRA)中扮演了重要角色。在 MRA 中,尺度函数被用于建立某一函数或图像的一系列近似值,相邻两近似值之间的近似度相差 2 倍。被称为小波的附加函数用于对相邻近似值之间的差异进行编码。

7.2.1 序列展开

信号或函数 $f(x)$ 常常可以被很好地分解为一系列展开函数的线性组合。

$$f(x) = \sum_k \alpha_k \varphi_k(x) \tag{7.2.1}$$

其中,k 是有限或无限和的整数下标,α_k 是具有实数值的展开系数,$\varphi_k(x)$ 是具有实数值的展开函数。如果展开是惟一的,也就是说对任何指定的 $f(x)$ 只有一个 α_k 系列与之相对应,则 $\varphi_k(x)$ 称为基函数,展开序列 $\{\varphi_k(x)\}$ 称为可被这样表示的一类函数的基。可展开的函数组成了一个函数空间,被称为展开集合的闭合跨度,表示为:

$$V = \overline{\mathrm{Span}_k\{\varphi_k(x)\}} \tag{7.2.2}$$

$f(x) \in V$ 表示 $f(x)$ 属于 $\{\varphi_k(x)\}$ 的闭合跨度,并能写成式(7.2.1)的形式。

对于任意函数空间 V 及其相应的展开集合 $\{\varphi_k(x)\}$ 都有一个二重函数集合,表示为 $\{\tilde{\varphi}_k(x)\}$,它可用于为任意 $f(x) \in V$ 计算式(7.2.1)的系数 α_k。这些系数可以通过计算 $\tilde{\varphi}_k(x)$ 和函数 $f(x)$ 的内积①得到。即:

$$\alpha_k = \langle \tilde{\varphi}_k(x), f(x) \rangle = \int \tilde{\varphi}_k^*(x) f(x) \mathrm{d}x \tag{7.2.3}$$

其中 * 表示复共轭操作。由于展开集合的正交性,该计算可以是 3 种可能形式中的一种。本章末的习题 7.10 使用二维欧氏空间的向量说明了这 3 种情况。

情况 1:如果该展开函数构成了 V 的一个正交基,即:

$$\langle \varphi_j(x), \varphi_k(x) \rangle = \delta_{jk} = \begin{cases} 0 & j \neq k \\ 1 & j = k \end{cases} \tag{7.2.4}$$

基与它的对偶相等。即,$\varphi_k(x) = \tilde{\varphi}_k(x)$,式(7.2.3)变成:

$$\alpha_k = \langle \varphi_k(x), f(x) \rangle \tag{7.2.5}$$

α_k 由基函数和 $f(x)$ 的内积来计算。

情况 2:如果展开函数本身不正交,而是 V 的正交基,那么:

$$\langle \varphi_j(x), \varphi_k(x) \rangle = 0 \quad , \quad j \neq k \tag{7.2.6}$$

基函数及其对偶称为双正交。使用式(7.2.3)计算 α_k,有:

---

① 两个实数或复数值函数 $f(x)$ 和 $g(x)$ 的内积是 $\langle f(x), g(x) \rangle = \int f^*(x) g(x) \mathrm{d}x$。如果 $f(x)$ 是实数,$f^*(x) = f(x)$ 和 $\langle f(x), g(x) \rangle = \int f(x) g(x) \mathrm{d}x$。

$$\langle \varphi_j(x), \tilde{\varphi}_k(x) \rangle = \delta_{jk} = \begin{cases} 0 & j \neq k \\ 1 & j = k \end{cases} \tag{7.2.7}$$

情况 3：如果展开集合对 V 来说不是函数基，但支持式(7.2.1)中定义的展开，那么它是一个跨度集合，对于任一 $f(x) \in V$ 有一个以上 α_k 集合。展开函数及其对偶称为超完备或冗余。它们组成了一个框架，其中①：

$$A \| f(x) \|^2 \leqslant \sum_k |\langle \varphi_k(x), f(x) \rangle|^2 \leqslant B \| f(x) \|^2 \tag{7.2.8}$$

对于某些 $A > 0, B < \infty$，及所有 $f(x) \in V$。用 $f(x)$ 的范数的平方去除该等式，可以看到 A, B"框架"构成了扩展系数以及函数的归一化内积。类似式(7.2.3)和式(7.2.5)的等式被用于寻找框架的展开系数。若 $A = B$，则展开集称为紧框架，它显示为如下形式(Daubechies[1992])：

$$f(x) = \frac{1}{A} \sum_k \langle \varphi_k(x), f(x) \rangle \varphi_k(x) \tag{7.2.9}$$

除 A^{-1} 项之外，它是框架冗余的度量，这与将式(7.2.5)(用于正交基)代入式(7.2.1)得到的表达式相等。

7.2.2　尺度函数

现在来考虑由整数平移和实数二值尺度、平方可积函数 $\varphi(x)$ 组成的展开函数集合，即集合 $\{\varphi_{j,k}(x)\}$，其中，

$$\varphi_{j,k}(x) = 2^{j/2} \varphi(2^j x - k) \tag{7.2.10}$$

对所有的 $j, k \in \mathbf{Z}$ 和 $\varphi(x) \in L^2(\mathbf{R})$ 都成立。② 此时，k 决定了 $\varphi_{j,k}(x)$ 在 x 轴上的位置，j 决定了 $\varphi_{j,k}(x)$ 的宽度，即沿 x 轴的宽或窄的程度，而 $2^{j/2}$ 控制其高度或幅度。由于 $\varphi_{j,k}(x)$ 的形状随 j 发生变化，$\varphi(x)$ 被称为尺度函数。通过选择适当的 $\varphi(x)$，$\{\varphi_{j,k}(x)\}$ 可以决定跨度 $L^2(\mathbf{R})$，所有可度量的平方可积函数的集合。

如果为式(7.2.10)中的 j 赋予一个定值，即 $j = j_0$，展开集合 $\{\varphi_{j_0,k}(x)\}$ 将是 $\varphi_{j,k}(x)$ 的一个子集。它并未跨整个 $L^2(\mathbf{R})$，而是其中的一个子空间。使用前面章节给出的符号，可将该子空间定义为：

$$V_{j_0} = \overline{\mathrm{Span}_k \{\varphi_{j_0,k}(x)\}} \tag{7.2.11}$$

也就是说，V_{j_0} 是 $\varphi_{j_0,k}(x)$ 在 k 上的一个跨度。如果 $f(x) \in V_{j_0}$，可写成：

$$f(x) = \sum_k \alpha_k \varphi_{j_0,k}(x) \tag{7.2.12}$$

更一般的情况下，定义下式代表对任何 j, k 上的跨度子空间：

$$V_j = \overline{\mathrm{Span}_k \{\varphi_{j,k}(x)\}} \tag{7.2.13}$$

如下例所示，增加 j 将增加 V_j 的大小，允许具有变化较小的变量或较细的细节函数包含在子空间中。这是由于 j 增大时，用于表示子空间函数的 $\varphi_{j,k}(x)$ 范围变窄，x 有较小变化即可分开。

---

① $f(x)$ 的范数，表示为 $\| f(x) \|$，定义为 $f(x)$ 与它本身内积的平方根。

② 注意 $L^2(\mathbf{R})$(这里 $\mathbf{R}$ 是实数集)表示为可度量集，整数平方的一维函数，$\mathbf{Z}$ 是整数集。

例 7.4 哈尔尺度函数

考虑单位高度、单位宽度的尺度函数(Haar[1910])：

$$\varphi(x) = \begin{cases} 1 & 0 \leqslant x < 1 \\ 0 & \text{其他} \end{cases} \tag{7.2.14}$$

图 7.9(a)到(d)显示了多个展开函数中的 4 个，这些展开函数通过将脉冲型尺度函数代入式(7.2.10)得到。注意，当 $j = 1$ 时，与 $j = 0$ 时恰好相反，展开函数更窄且更为密集。

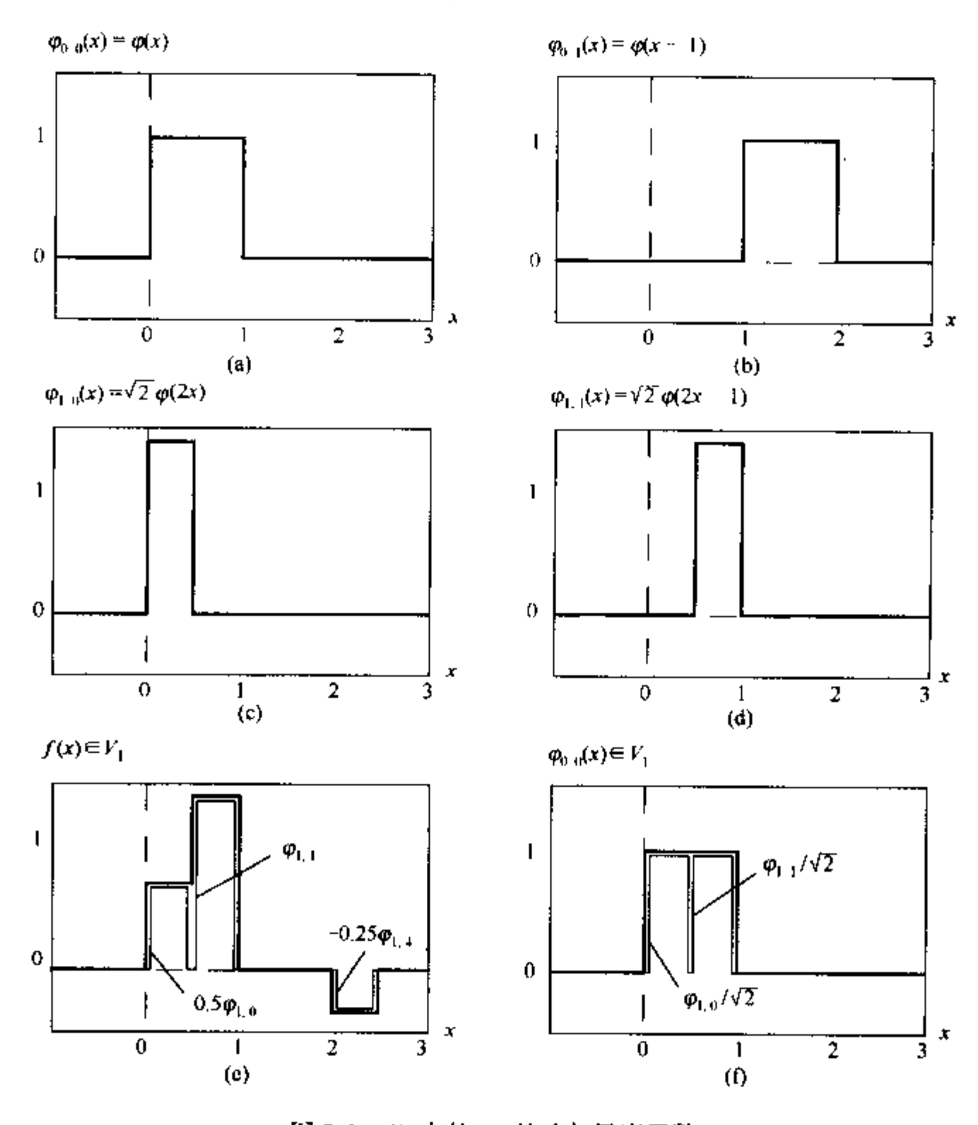


图 7.9 V_1 中的 V_0 的哈尔尺度函数

图 7.9(e)显示了子空间 V_1 中的一个成员。由于图 7.9(a)和(b)中的 V_0 的展开函数太粗糙而不能表示它，该函数不属于 V_0。需要如图 7.9(c)和(d)中的高分辨率函数。如图7.9(e)中所示，它们可以用三项展开来表示该函数：

$$f(x) = 0.5\varphi_{1,0}(x) + \varphi_{1,1}(x) - 0.25\varphi_{1,4}(x)$$

为总结上述例子，图 7.9(f)说明了将 $\varphi_{0,0}(x)$分解作为 V_1 展开函数的和。简而言之，V_0 的展开函数可以用下式分解：

$$\varphi_{0,k}(x) = \frac{1}{\sqrt{2}}\varphi_{1,2k}(x) + \frac{1}{\sqrt{2}}\varphi_{1,2k+1}(x)$$

因此,如果 $f(x)$ 是 V_0 的元素,它还是 V_1 的元素。这是由于 V_0 的所有展开函数都属于 V_1。从数学角度上看,V_0 是 V_1 的一个子空间,记为 $V_0 \subset V_1$。

在上边的例子中,简单的尺度函数遵循了多分辨率分析(Mallat[1989a])的 4 个基本要求:

MRA 要求 1:尺度函数对其积分变换是正交的。

很容易看出在哈尔函数的情况下,因为无论什么时候只要尺度函数的值是 1,其积分变换就是 0,所以二者的乘积是 0。哈尔函数是紧支撑的,即,除被称为支撑区的有限区间外,函数值都为 0。事实上,其支撑区是 1,半开区间[0,1)外的支撑区的值是 0。必须注意,当尺度函数的支撑区大于 1 时,积分变换正交的要求将很难满足。

MRA 要求 2:由低尺度的尺度函数跨越的子空间在低尺度处嵌套在由高尺度跨越的子空间内。

如图 7.10 所示,包含高分辨率函数的子空间必须同时包含所有低分辨率函数。也就是说,

$$V_{-\infty} \subset \cdots \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots \subset V_{\infty} \tag{7.2.15}$$

此外,这些子空间还满足直观条件,即,如果 $f(x) \in V_j$,那么 $f(2x) \in V_{j+1}$。哈尔尺度函数满足该要求并不意味着任何支撑区为 1 的函数都自动满足该条件。这作为一个练习留给读者去证明,同样简单的函数:

$$\varphi(x) = \begin{cases} 1 & 0.25 \leq x < 0.75 \\ 0 & 其他 \end{cases}$$

就不是多分辨率分析的有效尺度函数(见习题 7.11)。

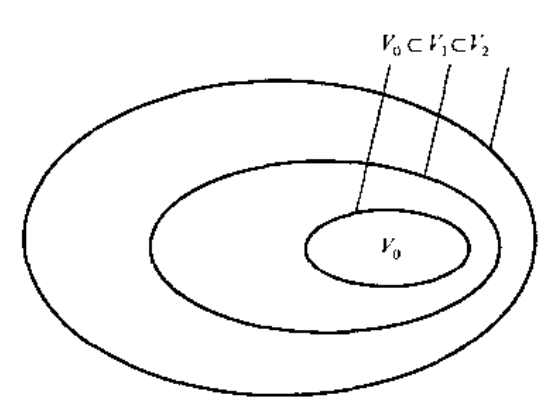


图 7.10　由尺度函数跨越的嵌套函数

MRA 要求 3:惟一包含在所有 V_j 中的函数是 $f(x)=0$。

如果考虑可能的最粗糙的展开函数(即 $j=-\infty$),惟一可表达的函数就是没有信息的函数,即,

$$V_{-\infty} = \{0\} \tag{7.2.16}$$

MRA 要求 4:任何函数都可以以任意精度表示。

虽然在任意粗糙的分辨率下展开一个特定 $f(x)$ 是几乎不可能的,像图 7.9(e)中所示的函数一样,但所有可度量的、平方可积函数都可以用极限 $j \to \infty$ 表示,即,

$$V_{\infty} = \{L^2(\mathbf{R})\} \tag{7.2.17}$$

在这些条件下,子空间 V_j 的展开函数可以被表述为子空间 V_{j+1} 的展开函数的加权和。使用式(7.2.12),令

$$\varphi_{j,k}(x) = \sum_n \alpha_n \varphi_{j+1,n}(x)$$

其中,求和的下标被改写成 n,以示区别。将式(7.2.10)代入 $\varphi_{j+1,n}(x)$,并将变量 α_n 改写成 $h_{\varphi}(n)$,上式变成:

$$\varphi_{j,k}(x) = \sum_n h_{\varphi}(n) 2^{(j+1)/2} \varphi(2^{j+1}x - n)$$

既然 $\varphi(x) = \varphi_{0,0}(x)$,$j$ 和 k 都可以置为 0,以得到较简单的无下标表达式:

$$\varphi(x) = \sum_n h_{\varphi}(n) \sqrt{2} \varphi(2x - n) \tag{7.2.18}$$

该递归等式中的系数 $h_{\varphi}(n)$ 被称为尺度函数系数;h_{φ} 为尺度矢量。其他文献中也常使用 $h(n)$ 和 $h_0(n)$,但我们希望能避免与前述子带分析滤波器发生混淆。式(7.2.18)是多分辨率分析的基础,称为改善等式、MRA 等式或扩张等式。它表示,任意子空间的展开函数都可以从它们自身的双倍分辨率拷贝中得到,即从相邻较高分辨率的空间中得到。对引用子空间 V_0 的选择是任意的。

例 7.5 哈尔尺度函数系数

对式(7.2.14)的哈尔函数的尺度函数系数是 $h_{\varphi}(0) = h_{\varphi}(1) = 1/\sqrt{2}$,即式(7.1.28)中矩阵 $\mathbf{H}_2$ 的第一行。因此,由式(7.2.18)可得:

$$\varphi(x) = \frac{1}{\sqrt{2}}[\sqrt{2}\varphi(2x)] + \frac{1}{\sqrt{2}}[\sqrt{2}\varphi(2x - 1)]$$

图 7.9(f)的 $\varphi_{0,0}(x)$ 很好地说明了这一分解过程,前边的表达式中用方括号括起来的项是 $\varphi_{1,0}(x)$ 和 $\varphi_{1,1}(x)$。附加的简化产生了 $\varphi(x) = \varphi(2x) + \varphi(2x - 1)$。

7.2.3 小波函数

给定满足上述 MRA 要求的尺度函数,能够定义小波函数 $\psi(x)$(与它的积分变换及其二进制尺度),跨越了相邻两尺度子空间 V_j 和 V_{j+1} 的差异。图 7.11 说明了这种情况。为图中跨越 W_j 空间的所有 $k \in \mathbf{Z}$ 定义小波集合 $\{\psi_{j,k}(x)\}$:

$$\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k) \tag{7.2.19}$$

使用尺度函数,可得:

$$W_j = \overline{\mathrm{Span}_k\{\psi_{j,k}(x)\}} \tag{7.2.20}$$

并注意,如果 $f(x) \in W_j$,

$$f(x) = \sum_k \alpha_k \psi_{j,k}(x) \tag{7.2.21}$$

尺度和图 7.11 中的小波函数子空间由下式相关联:

$$V_{j+1} = V_j \oplus W_j \tag{7.2.22}$$

这里,$\oplus$ 表示空间并集(类似于集合并集)。V_{j+1} 中 V_j 的正交补集是 W_j,且 V_j 中的所有成员对

于 W_j 中的所有成员都正交。因此，

$$\langle \varphi_{j,k}(x), \psi_{j,l}(x) \rangle = 0 \tag{7.2.23}$$

对所有适当的 $j,k,l \in \mathbf{Z}$ 都成立。

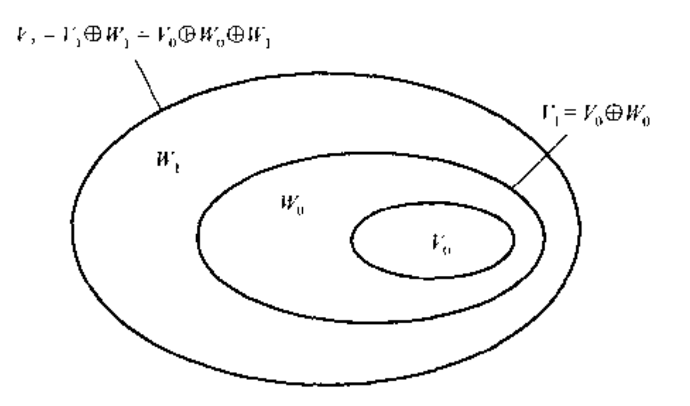


图 7.11　尺度及小波函数空间的关系

现在可以将所有可度量的、平方可积函数空间表示如下：

$$L^2(\mathbf{R}) = V_0 \oplus W_0 \oplus W_1 \oplus \ldots \tag{7.2.24}$$

$$L^2(\mathbf{R}) = V_1 \oplus W_1 \oplus W_2 \oplus \ldots \tag{7.2.25}$$

或者

$$L^2(\mathbf{R}) = \ldots \oplus W_{-2} \oplus W_{-1} \oplus W_0 \oplus W_1 \oplus W_2 \oplus \cdots, \tag{7.2.26}$$

上式中不出现尺度函数，函数仅用小波项进行表示。注意，如果 $f(x)$ 是 V_1 而不是 V_0 的元素，使用式(7.2.24)的展开式包含 $f(x)$ 使用 V_0 尺度函数的近似；来自 W_0 的小波将对近似与真实函数之间的差异进行编码。由式(7.2.24)到式(7.2.26)可得：

$$L^2(\mathbf{R}) = V_{j_0} \oplus W_{j_0} \oplus W_{j_0+1} \oplus \ldots \tag{7.2.27}$$

其中 j_0 是任意的开始尺度。

因为小波空间存在于由相邻较高分辨率尺度函数跨越的空间中(见图 7.11)，任何小波函数——类似式(7.2.18)中其尺度函数的对应部分——可以表示成平移的双倍分辨率尺度函数的加权和。可以写成：

$$\psi(x) = \sum_n h_\psi(n)\sqrt{2}\varphi(2x - n) \tag{7.2.28}$$

其中，$h_\psi(n)$ 称为小波函数系数，h_ψ 称为小波向量[文献中常用 $h_1(n)$ 表示]。利用小波跨越图 7.11中的正交补集空间且积分小波变换是正交的条件，可以显示 $h_\psi(n)$ 和 $h_\varphi(n)$ 以下述方式相关(见Burrus，Gopinath 和 Guo[1998]的例子)：

$$h_\psi(n) = (-1)^n h_\varphi(1 - n) \tag{7.2.29}$$

注意该结果与式(7.1.23)的相似性，该关系决定了正交子带编译码滤波器的冲激响应。

例 7.6　哈尔小波函数系数

在上例中，哈尔尺度向量定义为 $h_\varphi(0) = h_\varphi(1) = 1/\sqrt{2}$。使用式(7.2.29)，相应的小波向量为 $h_\psi(0) = (-1)^0 h_\varphi(1-0) = 1/\sqrt{2}$和 $h_\psi(1) = (-1)^1 h_\varphi(1-1) = -1/\sqrt{2}$。注意这些系数

与式(7.1.28)中矩阵 $\mathbf{H}_2$ 的第二行相对应。将这些值代入式(7.2.28),可得 $\psi(x)=\varphi(2x)-\varphi(2x-1)$,如图7.12(a)中所示。因此,哈尔小波函数为:

$$\psi(x)=\begin{cases}1 & 0\leqslant x<0.5\\ -1 & 0.5\leqslant x<1\\ 0 & \text{其他}\end{cases}\tag{7.2.30}$$

使用式(7.2.19),现在可产生尺度化且变换过的哈尔小波通式。图7.12(b)和(c)中分别显示了此类两个小波 $\psi_{0,2}(x)$ 和 $\psi_{1,0}(x)$。注意,对于空间 W_1,小波 $\psi_{1,0}(x)$ 比针对 W_0 的小波 $\psi_{0,2}(x)$ 窄,它可以用来表示更细微的细节。

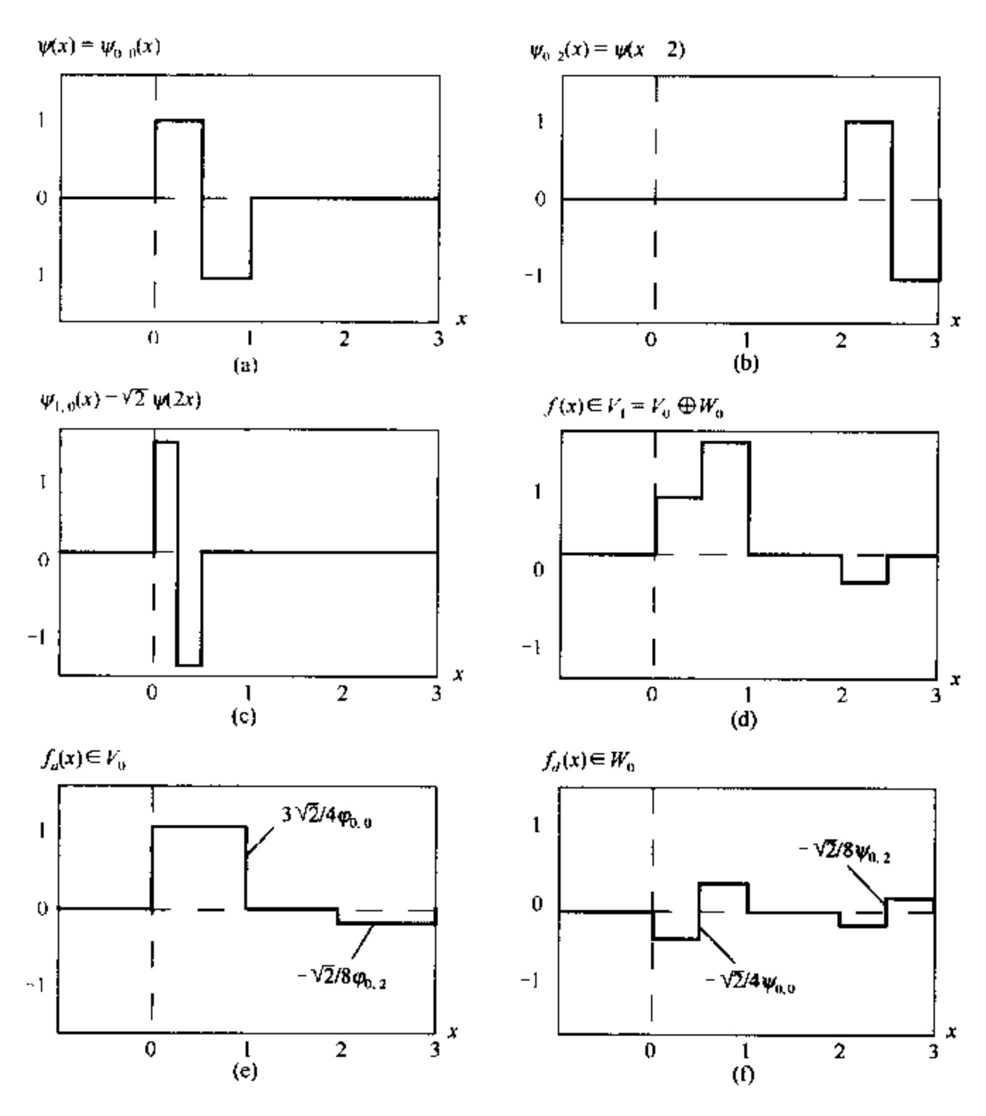


图7.12 在 W_0 和 W_1 中的哈尔小波函数

图7.12(d)显示了在子空间 V_1 而不在子空间 V_0 中的函数。该函数在前述例子中曾考虑过[见图7.9(e)]。虽然该函数不能在 V_0 中精确表示,式(7.2.22)指出,它可以用 V_0 和 W_0 的展开函数进行展开。展开结果如下:

$$f(x)=f_a(x)+f_d(x)$$

这里，

$$f_a(x) = \frac{3\sqrt{2}}{4}\varphi_{0,0}(x) - \frac{\sqrt{2}}{8}\varphi_{0,2}(x)$$

和

$$f_d(x) = \frac{-\sqrt{2}}{4}\psi_{0,0}(x) - \frac{\sqrt{2}}{8}\psi_{0,2}(x)$$

其中，$f_a(x)$是$f(x)$使用 V_0 尺度函数的近似，而 $f_d(x)$为 $f(x)-f_a(x)$的差，用 W_0 小波和表示。这两个展开式，如图 7.12(e)和(f)所示，将 $f(x)$用类似高通和低通滤波器的方法分成两部分。$f(x)$的低频部分在 $f_a(x)$中得到——$f_a(x)$给出了 $f(x)$在每个积分区间上的平均值——而高频细节则在 $f_d(x)$中编码。

7.3　一维小波变换

现在可以正式定义若干密切相关的小波变换：一般小波序列展开、离散小波变换和连续小波变换。它们在傅里叶域的对应部分分别是傅里叶序列展开、离散傅里叶变换和连续傅里叶变换。在 7.4 节，将定义一种计算效率很高的称做快速小波变换的离散小波变换。

7.3.1　小波序列展开

首先根据小波 $\psi(x)$和尺度函数 $\varphi(x)$为函数 $f(x)\in L^2(\mathbf{R})$定义小波序列展开。据式(7.2.27)，可写出：

$$f(x) = \sum_k c_{j_0}(k)\varphi_{j_0,k}(x) + \sum_{j=j_0}^{\infty}\sum_k d_j(k)\psi_{j,k}(x) \tag{7.3.1}$$

其中 j_0 是任意开始尺度，c_{j_0} 和 $d_j(k)$分别是式(7.2.12)和式(7.2.21)中 α_k 的改写。$c_{j_0}(k)$通常称为近似值或尺度系数；$d_j(k)$称为细节或小波系数。这是因为式(7.3.1)的第一个和式用尺度函数提供了 $f(x)$在尺度 j_0 的近似[除非 $f(x)\in V_{j_0}$，此时为其精确值]。对于第二个和式中每一个较高尺度的 $j\geqslant j_0$，更细分辨率的函数(一个小波和)被添加到近似中以获得细节的增加。如果展开函数形成了一个正交基或紧框架(通常情况下是这样)，基于式(7.2.5)和式(7.2.9)的展开系数计算如下：

$$c_{j_0}(k) = \langle f(x), \varphi_{j_0,k}(x)\rangle = \int f(x)\varphi_{j_0,k}(x)\mathrm{d}x \tag{7.3.2}$$

和

$$d_j(k) = \langle f(x), \psi_{j,k}(x)\rangle = \int f(x)\psi_{j,k}(x)\mathrm{d}x \tag{7.3.3}$$

如果展开函数是双正交基的一部分，上式中的 φ 和 ψ 项要分别由它们的对偶函数和代替。

例 7.7　$y=x^2$ 的哈尔小波序列展开

考虑图 7.13(a)中显示的简单函数：

$$y = \begin{cases} x^2 & 0\leqslant x<1 \\ 0 & \text{其他} \end{cases}$$

使用哈尔小波——见式(7.2.14)和式(7.2.30)——和开始尺度 $j_0=0$，式(7.3.2)和

式(7.3.3)可以被用来计算下述展开系数：

$$c_0(0)=\int_0^1 x^2\varphi_{0,0}(x)\mathrm{d}x=\int_0^1 x^2\mathrm{d}x=\frac{x^3}{3}\Big|_0^1=\frac{1}{3}$$

$$d_0(0)=\int_0^1 x^2\psi_{0,0}(x)\mathrm{d}x=\int_0^{0.5}x^2\mathrm{d}x-\int_{0.5}^1 x^2\mathrm{d}x=-\frac{1}{4}$$

$$d_1(0)=\int_0^1 x^2\psi_{1,0}(x)\mathrm{d}x=\int_0^{0.25}x^2\sqrt{2}\mathrm{d}x-\int_{0.25}^{0.5}x^2\sqrt{2}\mathrm{d}x=-\frac{\sqrt{2}}{32}$$

$$d_1(1)=\int_0^1 x^2\psi_{1,1}(x)\mathrm{d}x=\int_0^{0.75}x^2\sqrt{2}\mathrm{d}x-\int_{0.75}^{1}x^2\sqrt{2}\mathrm{d}x=-\frac{3\sqrt{2}}{32}$$

图 7.13　使用哈尔小波，$y=x^2$ 的小波序列展开

将这些值代入式(7.3.1)，可以得到小波序列展开：

$$y=\underbrace{\underbrace{\underbrace{\frac{1}{3}\varphi_{0,0}(x)}_{V_0}+\underbrace{\left[-\frac{1}{4}\psi_{0,0}(x)\right]}_{W_0}}_{V_1=V_0\oplus W_0}+\underbrace{\left[-\frac{\sqrt{2}}{32}\psi_{1,0}(x)-\frac{3\sqrt{2}}{32}\psi_{1,1}(x)\right]}_{W_1}}_{V_2=V_1\oplus W_1=V_0\oplus W_0\oplus W_1}+\cdots$$

上述展开中的第一项用 $c_0(0)$ 生成待展开函数的 V_0 子空间近似值。该近似值如

图7.13(b)所示,是原始函数的平均值。第二项使用 $d_0(0)$ 通过从 W_0 子空间添加一级细节来修饰该近似值。添加的细节及 V_1 的结果近似值分别如图7.13(c)和(d)所示。其他级别的细节由子空间 W_1 的系数 $d_1(0)$ 和 $d_1(1)$ 给出。该附加细节如图7.13(e)所示,V_2 的结果近似值如图7.13(f)所示。注意,展开函数现在已经接近原始函数了。越高的尺度(或细节的级别越高)被叠加,近似值越变得接近函数的精确表示,它的极限是 $j\to\infty$。

7.3.2 离散小波变换

与傅里叶序列展开相似,前节的小波序列展开将一个连续变量函数映射成一系数序列。如果待展开函数是一个数字序列,如连续函数 $f(x)$ 的抽样值,得到的系数就称为 $f(x)$ 的离散小波变换(DWT)。在这种情况下,式(7.3.1)到式(7.3.3)中定义的序列展开变成DWT变换对:

$$W_\varphi(j_0,k)=\frac{1}{\sqrt{M}}\sum_x f(x)\varphi_{j_0,k}(x) \tag{7.3.5}$$

$$W_\psi(j,k)=\frac{1}{\sqrt{M}}\sum_x f(x)\psi_{j,k}(x) \tag{7.3.6}$$

对于 $j\geq j_0$ 还有:

$$f(x)=\frac{1}{\sqrt{M}}\sum_k W_\varphi(j_0,k)\varphi_{j_0,k}(x)+\frac{1}{\sqrt{M}}\sum_{j=j_0}^{\infty}\sum_k W_\psi(j,k)\psi_{j,k}(x) \tag{7.3.7}$$

这里,$f(x)$,$\varphi_{j_0,k}(x)$ 和 $\psi_{j,k}(x)$ 是离散变量 $x=0,1,2,\cdots,M-1$ 的函数。例如,对于某个 x_0,Δx,$x=0,1,2,\cdots,M-1$,$f(x)=f(x_0+x\Delta x)$。通常,令 $j_0=0$ 并选择 M 是2的幂(即 $M=2^J$),对 $x=0,1,2,\cdots,M-1$,$j=0,1,2,\cdots,J-1$ 和 $k=0,1,2,\cdots,2^j-1$ 求和。对于哈尔小波,变换中使用的离散尺度和小波函数(即基函数)与7.1.3节中的 $M\times M$ 哈尔变换矩阵的行相对应。变换本身由 M 个系数组成,最小尺度是0,最大尺度是 $J-1$。出于7.3.1节和例7.6中解释的原因,式(7.3.5)和式(7.3.6)中定义的系数分别称为近似值和细节系数。

式(7.3.5)到式(7.3.7)中的 $W_\varphi(j_0,k)$ 和 $W_\psi(j,k)$ 对应于前节的小波序列展开中的 $c_{j_0}(k)$ 和 $d_j(k)$(这些变量的变化并不是必须的,但它为下一节的连续小波变换中的标准符号提供了铺垫)。注意,序列展开中的积分变成了求和,而曾在4.2.1节的DFT中出现的归一化因子 $1/\sqrt{M}$ 在展开和反展开表达式中都有出现。该因子也可以在正展开和反展开表达式中以 $1/M$ 的形式出现。最后,应该记住,式(7.3.5)到式(7.3.7)只对正交基和紧框架有效。对于双正交基,式(7.3.5)和式(7.3.6)中的 φ 和 ψ 项必须由它们的对偶函数 $\tilde{\varphi}$ 和 $\tilde{\psi}$ 代替。

例7.8　计算一维离散小波变换

为说明式(7.3.5)到式(7.3.7)的用途,考虑四点的离散函数:$f(0)=1$,$f(1)=4$,$f(2)=-3$ 和 $f(3)=0$。因为 $M=4$,$J=2$ 且由于 $j_0=0$,对 $x=0,1,2,3$,$j=0,1$(且对于 $j=0$,$k=0$,或者对于 $j=1$,$k=0,1$)求和。将使用哈尔尺度和小波函数,并假定 $f(x)$ 的4个采样值分布在基函数的支撑区上,基函数的值为1。将4个采样点代入式(7.3.5),可得:

$$\begin{aligned}W_\varphi(0,0)&=\frac{1}{2}\sum_{x=0}^{3}f(x)\varphi_{0,0}(x)\\&=\frac{1}{2}[1\cdot1+4\cdot1-3\cdot1+0\cdot1]=1\end{aligned}$$

因为对于 $x=0,1,2,3,\varphi_{0,0}(x)=1$。注意,这里采用的是哈尔尺度函数对于 $j=0$ 且 $k=0$ 的均匀空间采样。该值对应于 7.1.3 节的哈尔变换矩阵 $\mathbf{H}_4$ 的第一行。继续使用式(7.3.6)和相似间隔的采样点 $\psi_{j,k}(x)$(它对应于 $\mathbf{H}_4$ 的第 2,3,4 行),可得:

$$W_\psi(0,0)=\frac{1}{2}[1\cdot 1+4\cdot 1-3\cdot(-1)+0\cdot(-1)]=4$$

$$W_\psi(1,0)=\frac{1}{2}[1\cdot\sqrt{2}+4\cdot(-\sqrt{2})-3\cdot 0+0\cdot 0]=-1.5\sqrt{2}$$

$$W_\psi(1,1)=\frac{1}{2}[1\cdot 0+4\cdot 0-3\cdot\sqrt{2}+0\cdot(-\sqrt{2})]=-1.5\sqrt{2}$$

因此,这种简单的四点采样函数的离散小波变换与哈尔小波和尺度函数的关系是 $\{1,4,-1.5\sqrt{2},-1.5\sqrt{2}\}$。这里,变换系数是以它们计算的顺序安排的。

式(7.3.7)允许从变换中恢复出原始函数。重复求和,可得:

$$f(x)=\frac{1}{2}[W_\varphi(0,0)\varphi_{0,0}(x)+W_\psi(0,0)\psi_{0,0}(x)+W_\psi(1,0)\psi_{1,0}(x)+W_\psi(1,1)\psi_{1,1}(x)]$$

$x=0,1,2,3$。如果 $x=0$,

$$f(0)=\frac{1}{2}[1\cdot 1+4\cdot 1-1.5\sqrt{2}\cdot(\sqrt{2})-1.5\sqrt{2}\cdot 0]=1$$

与正变换情况一样,尺度和小波函数的均匀空间采样也用于反变换的计算。

上例中的 4 点 DWT 是 $f(x)$ 的一个二尺度分解,即 $j=\{0,1\}$。基本的假设是开始尺度 j_0 为0,也可以使用其他开始尺度。留给读者作为练习计算单尺度变换 $\{2.5\sqrt{2},-1.5\sqrt{2},-1.5\sqrt{2},-1.5\sqrt{2}\}$,当开始尺度是 1 时,计算其结果(见习题 7.16)。这样,式(7.3.5)和式(7.3.6)定义了一组开始尺度与 j_0 不同的变换。

7.3.3 连续小波变换

离散小波变换的自然延伸是连续小波变换(CWT),将一个连续函数变换成两个连续变量(变换和尺度)的高冗余度函数。变换结果在时频分析上很容易解释并有很大价值。虽然我们的兴趣在离散图像上,出于完整性的考虑,也覆盖了连续变换。

连续的平方可积函数 $f(x)$ 的连续小波变换与实数值的小波 $\psi(x)$ 的关系如下:

$$W_\psi(s,\tau)=\int_{-\infty}^{\infty}f(x)\psi_{s,\tau}(x)\mathrm{d}x \tag{7.3.8}$$

其中,

$$\psi_{s,\tau}(x)=\frac{1}{\sqrt{s}}\psi\left(\frac{x-\tau}{s}\right) \tag{7.3.9}$$

s 和 τ 分别称为尺度和变换参数。给定 $W_\psi(s,\tau)$,可以通过反连续小波变换求得 $f(x)$:

$$f(x)=\frac{1}{C_\psi}\int_0^{\infty}\int_{-\infty}^{\infty}W_\psi(s,\tau)\frac{\psi_{s,\tau}(x)}{s^2}\mathrm{d}\tau\mathrm{d}s \tag{7.3.10}$$

其中

$$C_\psi=\int_{-\infty}^{\infty}\frac{|\Psi(u)|^2}{|u|}\mathrm{d}u \tag{7.3.11}$$

$\Psi(u)$是 $\psi(x)$的傅里叶变换。式(7.3.8)到式(7.3.11)定义了可反变换的变换，只要满足所谓的“允许条件”，$C_\psi < \infty$（Grossman和Morlet[1984]）。大多数情况下，这表示 $\Psi(0)=0$ 且 $u \to \infty$ 时 $\Psi(u) \to 0$，速度足够快以使 $C_\psi < \infty$。

上式是它们的离散对应部分[式(7.2.19)、式(7.3.1)、式(7.3.3)、式(7.3.6)和式(7.3.7)]的再现。注意下列相似性：

1. 连续变换参数τ取代了积分变换参数 k；
2. 连续尺度参数 s 与二进制尺度参数 2^j 相反。这是由于 s 出现在式(7.3.9)中 $\psi((x-\tau)/s)$的分母上。这样，连续变换所用的小波在 $0<s<1$ 时被压缩或宽度减小，在 $s>1$ 时扩大或展开。小波尺度和通常意义上的频率定义是相反的。
3. 连续变换与序列展开[见式(7.3.1)]或离散变换[见式(7.3.6)] 相似，其中开始尺度 $j_0=-\infty$。根据式(7.2.26)，这消除了尺度函数间的明显关联，因此，函数表示只包括小波项。
4. 与离散变换类似，连续变换可以被看成是一组变换系数$\{W_\psi(s,\tau)\}$，它给出了 $f(x)$与基函数集$\{\psi_{s,\tau}(x)\}$的相似性。然而，在连续的情况下，两个集合都是无穷的。由于 $\psi_{s,\tau}(x)$具有实数值且 $\psi_{s,\tau}(x)=\psi^*_{s,\tau}(x)$，式(7.3.8)中的每个系数都是 $f(x)$和 $\psi_{s,\tau}(x)$的整数内积$\langle f(x),\psi_{s,\tau}(x)\rangle$。

例 7.9　一维连续小波变换

墨西哥草帽小波：

$$\psi(x)=\left(\frac{2}{\sqrt{3}}\pi^{-1/4}\right)(1-x^2)e^{-x^2/2} \tag{7.3.12}$$

由于波形而得名[见图 7.14(a)]。它与高斯概率函数的二阶导数成比例，均值为零，是紧支撑的($|x|\to\infty$ 时迅速衰减)。虽然满足连续可逆变换存在的容许性要求，但没有相关的尺度函数，且变换不能产生正交分析。它最显著的特性是它的对称性和式(7.3.12)的显式表达的存在性。

图 7.14(a)中的连续一维函数是两个墨西哥草帽小波的和：

$$f(x)=\psi_{1,10}(x)+\psi_{6,80}(x)$$

其傅里叶频谱，如图 7.14(b)所示，揭示了尺度化的小波和傅里叶频段之间的联系。频谱中的两个显著频段(或峰值)对应函数的两个类高斯扰动。

图 7.14(c)显示了图 7.14(a)中的函数根据墨西哥草帽小波完成的 CWT 的一部分($1\le s\le 10$ 且$\tau\le 100$)。与图 7.14(b)中的傅里叶频谱不同，它同时给出了时域和频域的信息。注意，例如，当 $s=1$ 时，变换在$\tau=10$ 时达到最大值，对应 $f(x)$分量 $\psi_{1,10}(x)$的位置。因为变换为 $f(x)$及其计算出的小波之间的相似性提供了客观评价，很容易了解它是如何用于特性检测的。这里只需要与我们感兴趣的特性相匹配的小波。从图 7.14(d)的灰度图也可以得出类似结论，其中，变换的绝对值$|W_\psi(s,\tau)|$用黑白之间的灰度级显示。注意连续小波变换将一个一维函数变换成一个二维函数。

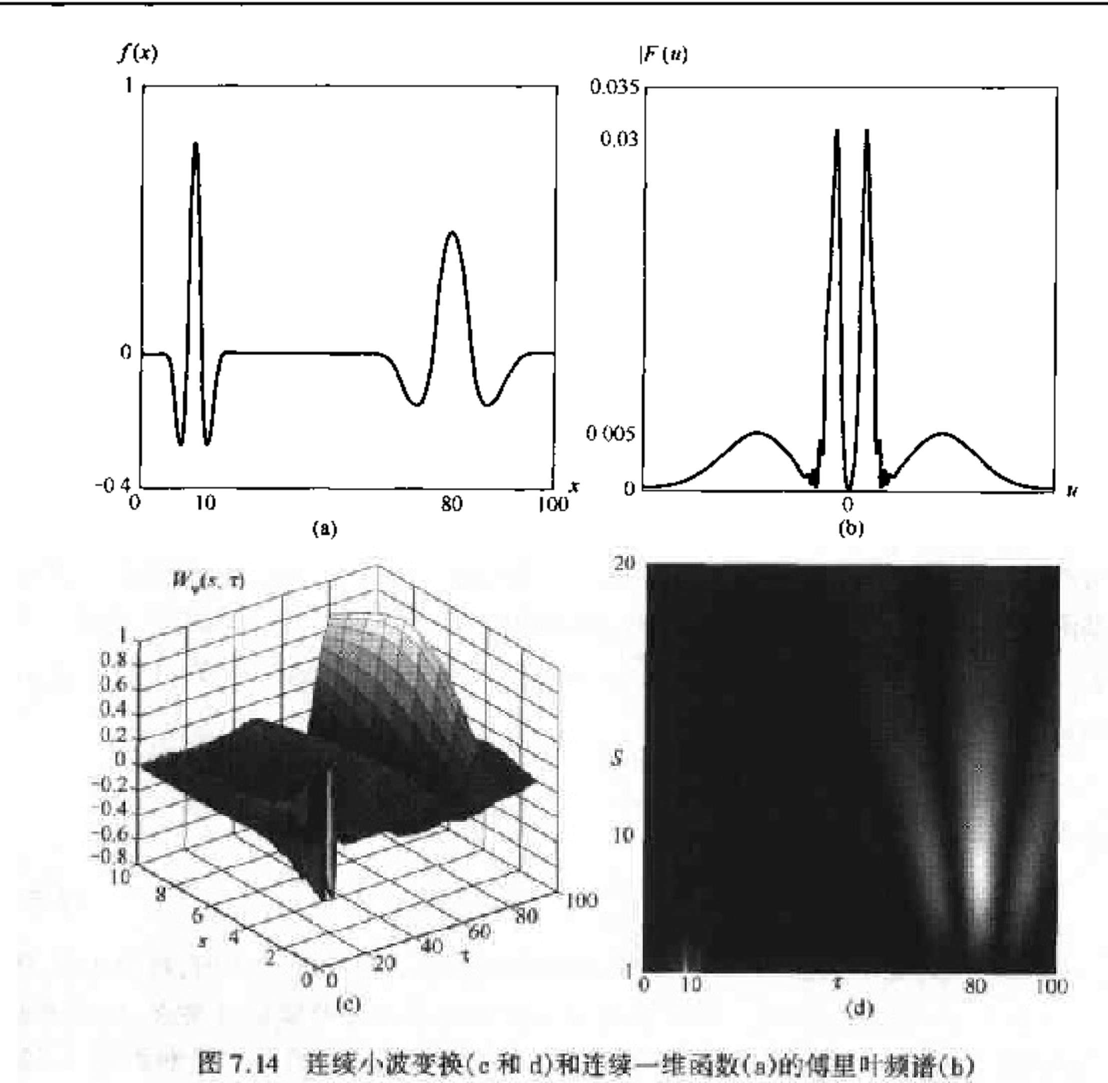


图 7.14 连续小波变换(c 和 d)和连续一维函数(a)的傅里叶频谱(b)

7.4 快速小波变换

快速小波变换(FWT)是一种实现离散小波变换(DWT)的高效计算,该变换找到了相邻尺度 DWT 系数间的一种令人惊喜的关系。它也称为Mallat人字形算法(Mallat[1989a,b]),FWT 类似于 7.1.2 节中的两段子带编码方案。

再次考虑多分辨率改善等式:

$$\varphi(x) = \sum_n h_\varphi(n)\sqrt{2}\varphi(2x - n) \tag{7.4.1}$$

用 2^j 对 x 进行尺度化,用 k 对它进行平移,令 $m = 2k + n$,得:

$$\begin{aligned}\varphi(2^j x - k) &= \sum_n h_\varphi(n)\sqrt{2}\varphi(2(2^j x - k) - n) \\ &= \sum_m h_\varphi(m - 2k)\sqrt{2}\varphi(2^{j+1}x - m)\end{aligned} \tag{7.4.2}$$

注意,尺度向量 h_φ 可以被看成是用来将 $\varphi(2^j x - k)$展开成尺度为 $j+1$ 的尺度函数和的"权"。从式(7.2.28)开始的一系列简单操作为 $\psi(2^j x - k)$得出类似结论。即,

$$\psi(2^j x - k) = \sum_m h_\psi(m - 2k)\sqrt{2}\varphi(2^{j+1}x - m) \tag{7.4.3}$$

其中,式(7.4.2)中的尺度向量 $h_\varphi(n)$ 被式(7.4.3)中的小波向量 $h_\psi(n)$ 取代。

现在考虑 7.2.2 节中的式(7.3.5)和式(7.3.6)。它们定义了离散小波变换。将小波定义式(7.2.19)代入式(7.3.6),可得:

$$W_\psi(j,k) = \frac{1}{\sqrt{M}}\sum_x f(x)2^{j/2}\psi(2^j x - k) \tag{7.4.4}$$

用式(7.4.3)的右端代替 $\psi(2^j x - k)$,变成:

$$W_\psi(j,k) = \frac{1}{\sqrt{M}}\sum_x f(x)2^{j/2}\Big[\sum_m h_\psi(m - 2k)\sqrt{2}\varphi(2^{j+1}x - m)\Big] \tag{7.4.5}$$

交换和式与整数,并重新安排,给出:

$$W_\psi(j,k) = \sum_m h_\psi(m - 2k)\Big[\frac{1}{\sqrt{M}}\sum_x f(x)2^{(j+1)/2}\varphi(2^{j+1}x - m)\Big] \tag{7.4.6}$$

这里,括起来的量与式(7.3.5)在 $j_0 = j + 1$ 时相等。将式(7.2.10)代入式(7.3.5)并令 $j_0 = j + 1$,这样可以写为:

$$W_\psi(j,k) = \sum_m h_\psi(m - 2k)W_\varphi(j + 1, m) \tag{7.4.7}$$

注意,DWT 在尺度 j 的细节系数是 DWT 在尺度 $j+1$ 的近似值系数的函数。将式(7.4.2)和式(7.3.5)作为包含DWT近似值系数的近似派生的起点,类似地,有:

$$W_\varphi(j,k) = \sum_m h_\varphi(m - 2k)W_\varphi(j + 1, m) \tag{7.4.8}$$

式(7.4.7)和式(7.4.8)揭示了 DWT 相邻尺度系数间的重要关系。比较上述结果和式(7.1.7),我们看到尺度 j 的近似值和细节系数 $W_\varphi(j,k)$ 与 $W_\psi(j,k)$ 可以通过做尺度 $j+1$ 的近似值系数 $W_\varphi(j+1,k)$ 和时域反转的尺度与小波向量[$h_\varphi(-n)$ 和 $h_\psi(-n)$]的卷积,而后对结果亚取样来计算。图 7.15 将这些操作简化成框图形式。注意,这与图 7.4 中的两段子带编译码系统的分析部分相同,$h_0(n) = h_\varphi(-n)$ 且 $h_1(n) = h_\psi(-n)$。因此,有:

$$W_\psi(j,k) = h_\psi(-n) * W_\varphi(j + 1, n)\big|_{n=2k,k\geq0} \tag{7.4.9}$$

和

$$W_\varphi(j,k) = h_\varphi(-n) * W_\varphi(j + 1, n)\big|_{n=2k,k\geq0} \tag{7.4.10}$$

其中,卷积在 $n = 2k$ 时进行计算($k \geq 0$)。在非负偶数时刻计算卷积与以 2 为步长进行过滤和抽样的效果相同。

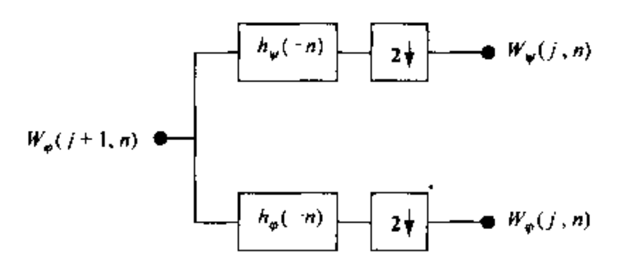


图 7.15　一个 FWT 分析滤波器族

为了总结 FWT 的发展,只需注意图 7.15 中的滤波器族可以迭代产生多阶结构,用于计算两个以上连续尺度的DWT系数。例如,图 7.16(a)显示了一个用于计算变换的两个最高尺度系数的二阶滤波器族。注意,最高的尺度系数假定是函数自身的采样值。即,$W_\varphi(J,n)=f(n)$,这里,J 表示最高的尺度[根据 7.2.2 节,$f(x)\in V_J$,V_J 是函数 $f(x)$所在的尺度空间]。图 7.16(a)中的第一个滤波器族将原始函数分解成一个低通近似值分量[其对应于尺度系数 $W_\varphi(J-1,n)$]和一个高通细节分量[其对应于系数 $W_\psi(J-1,n)$]。如图 7.16(b)所示,尺度空间 V_J 被分成小波子空间 W_{J-1}和尺度子空间 V_{J-1}。原始函数的频谱被分成两个半波段分量。图 7.16(a)中的第二个滤波器族将频谱和子空间 V_{J-1}(较低的半波段)分成四分之一波段子空间 W_{J-2}和 V_{J-2},分别对应于 DWT 系数 $W_\psi(J-1,n)$和 $W_\varphi(J-1,n)$。

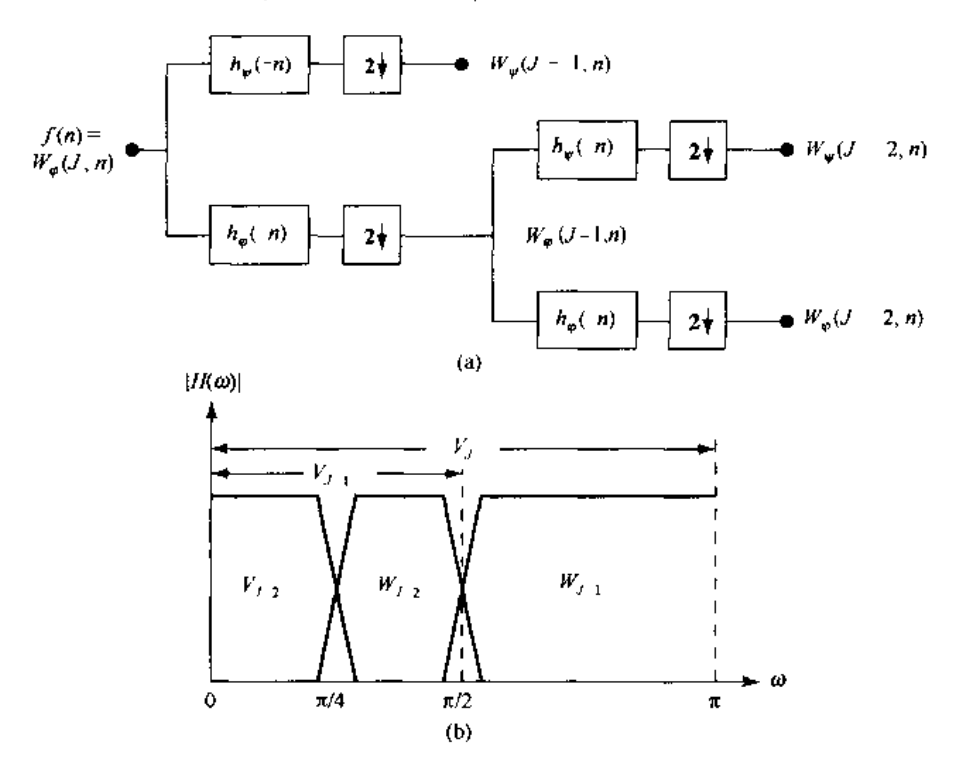


图 7.16 (a)一个两阶或两尺度 FWT 分析滤波器族,(b)它的频率分离特性

图 7.16 中的二阶滤波器族很容易进一步扩展得到任意阶数的滤波器族。例如,第三个滤波器族要处理 $W_\varphi(J-2,n)$系数,将尺度空间 V_{J-2}分成两个八分之一波段子空间 W_{J-3}和 V_{J-3}。通常,选择 $f(x)$的 2^J 个采样值,用 P 个滤波器族生成尺度 $J-1,J-2,\cdots,J-P$ 的 P 尺度 FWT。首先计算最高的尺度系数(即 $J-1$);最后计算最低的(即 $J-P$)。如果 $f(x)$以高于奈奎斯特频率的采样率进行采样(通常如此),其采样值是该采样分辨率的尺度系数的良好近似,并可以用做起始的高分辨率尺度系数的输入。换句话说,在该采样尺度,不需要小波或细节系数。分辨率最高的尺度函数作为式(7.3.5)和式(7.3.6)的德尔塔函数,允许 $f(n)$用做尺度 J 的近似值或尺度系数,输入到第一个两频段滤波器族(Odegard, Gopinath 和 Burrus [1992])。

例 7.10 计算一维小波变换

为解释这些概念,考虑例 7.8 中的离散函数 $f(n)=\{1,4,-3,0\}$。如例 7.8 所示,我们再

次计算基于哈尔尺度和小波函数的变换。然而，不再像例 7.8 中的DWT那样直接使用基函数，而是使用例 7.5 和例 7.6 中相应的尺度和小波向量：

$$h_{\varphi}(n)=\begin{cases}1/\sqrt{2} & n=0,1\\ 0 & \text{其他}\end{cases} \tag{7.4.11}$$

和

$$h_{\psi}(n)=\begin{cases}1\sqrt{2} & n=0\\ -1/\sqrt{2} & n=1\\ 0 & \text{其他}\end{cases} \tag{7.4.12}$$

这些是用于建立 FWT 滤波器族的函数；它们给出了滤波器系数。

既然例 7.8 计算的 DWT 由元素 $\{W_{\varphi}(0,0), W_{\psi}(0,0), W_{\psi}(1,0), W_{\psi}(1,1)\}$ 组成，将对尺度 $j=\{0,1\}$ 计算相应的二尺度 FWT。即，$J=2$（有 $2^J=2^2$ 个采样值）且 $P=2$（按尺度 $J-1=2-1=1$ 和 $J-P=2-2=0$ 的顺序进行）。变换使用图 7.16(a) 中的二阶滤波器族进行计算。图 7.17 显示了从要求的 FWT 卷积和抽样获得的序列。注意，函数 $f(n)$ 自身是最左边的滤波器族的尺度或近似值输入。例如，为计算出现在图 7.17 中上支路末端的 $W_{\psi}(1,k)$ 系数，首先要做 $f(n)$ 和 $h_{\psi}(-n)$ 的卷积。如 4.6.3 节所述，这样做要求关于原点反转某一原始函数，在另一函数上平移，逐点计算两函数各对应点的乘积和。对于序列 $\{1,4,-3,0\}$ 和 $\{-1/\sqrt{2}, 1/\sqrt{2}\}$，该结果为 $\{-1\sqrt{2}, -3/\sqrt{2}, 7/\sqrt{2}-3/\sqrt{2}, 0\}$，其中第二项对应下标 $k=2n=0$（在图 7.17 中，带下划线的值表示负时刻，即 $n<0$）。对偶数下标的点进行抽样，得到 $W_{\psi}(1,k)=\{-3/\sqrt{2}, -3/\sqrt{2}\}$ $(k=\{0,1\})$。也可以交替地使用式(7.4.9)计算：

$$\begin{aligned}W_{\psi}(1,k) &= h_{\psi}(-n) * W_{\varphi}(2,n)\big|_{n=2k,k\geqslant 0} = h_{\psi}(-n) * f(n)\big|_{n=2k,k\geqslant 0}\\ &= \sum_{l} h_{\psi}(2k-l)x(l)\Big|_{k=0,1}\\ &= \frac{1}{\sqrt{2}}x(2k) - \frac{1}{\sqrt{2}}x(2k+1)\Big|_{k=0,1}\end{aligned}$$

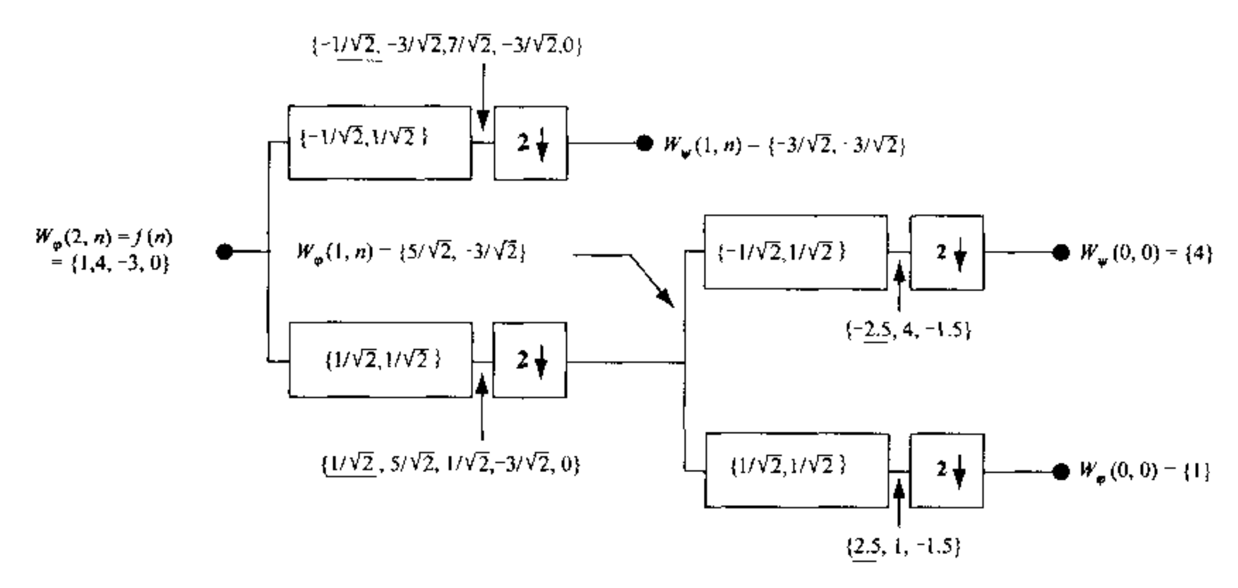


图 7.17　用哈尔尺度和小波向量计算序列 $\{1,4,-3,0\}$ 的二尺度快速小波变换

这里,用 $2k$ 代替卷积中的 n,用 l 作为卷积中的虚变量(即,替代相关的两序列)。展开和中只有两项,因为在小波向量 $h_\psi(-n)$ 的时域反转中只有两个非零项。令 $k=0$,得出 $W_\psi(1,0)=-3/\sqrt{2}$;对于 $k=1$,得到 $W_\psi(1,1)=-3/\sqrt{2}$。这样,滤波并抽样后的序列为 $\{-3/\sqrt{2},-3/\sqrt{2}\}$,与先前的结果匹配。同理可做其余的卷积和抽样。

从 DWT/FWT 的近似值和细节系数 $W_\varphi(j,k)$ 与 $W_\psi(j,k)$ 重建 $f(x)$ 也存在一种高效的反变换,称为快速小波反变换(FWT^{-1})。它使用正变换中所用的尺度和小波向量以及第 j 级近似值和细节系数来生成第 $j+1$ 级近似值系数。注意到图 7.15 中 FWT 分析部分和图 7.4(a)中两频段子带分析部分的相似性,可以立即得出要求的 FWT^{-1} 的综合滤波器族。图 7.18 详细描述了它的结构,与图 7.4(a)中两频段子带编译码系统的综合部分相同。7.1.2 节中的式(7.1.23)定义了相关的综合滤波器。与该处所述相同,完美重建(对于双子带或正交滤波器)要求对于 $i=\{0,1\}$,$g_i(n)=h_i(-n)$。也就是说,分析和综合滤波器在时域中是相互反转的。因为 FWT 的分析滤波器(见图 7.15)是 $h_0(n)=h_\varphi(-n)$ 和 $h_1(n)=h_\psi(-n)$,要求的 FWT^{-1} 的综合滤波器为 $g_0(n)=h_0(-n)=h_\varphi(n)$ 和 $g_1(n)=h_1(-n)=h_\psi(n)$。然而,应记住,也可能使用双正交分析和综合滤波器,它们并不是彼此时域反转的。双正交分析和综合滤波器是根据式(7.1.14)和式(7.1.15)交叉调制的。

图 7.18 中的 FWT^{-1} 的滤波器族执行下述计算:

$$W_\varphi(j+1,k)=h_\varphi(k)*W_\varphi^{\text{up}}(j,k)+h_\psi(k)*W_\psi^{\text{up}}(j,k)\big|_{k\geq 0} \tag{7.4.13}$$

其中 W^{up} 代表以 2 为步长进行内插(即,在 W 的各元素间插入 0,使其长度变为原来的两倍)。内插后的系数通过与 $h_\varphi(n)$ 和 $h_\psi(n)$ 进行卷积完成过滤,并相加以得到较高尺度的近似值。最终将建立 $f(x)$ 的较好近似,该近似含有较多的细节和较高的分辨率。与 FWT 正变换类似,反变换滤波器族可以如图 7.19 所示进行迭代,在这里,为了计算 FWT^{-1} 重建的最后两个尺度描绘了两尺度结构。该系数合并过程可以扩展到任意数目的尺度,从而保证函数 $f(x)$ 的完美重建。

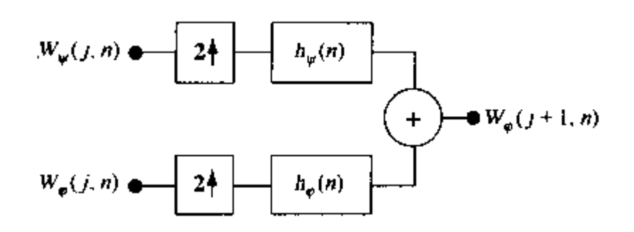


图 7.18 FWT^{-1} 的综合滤波器族

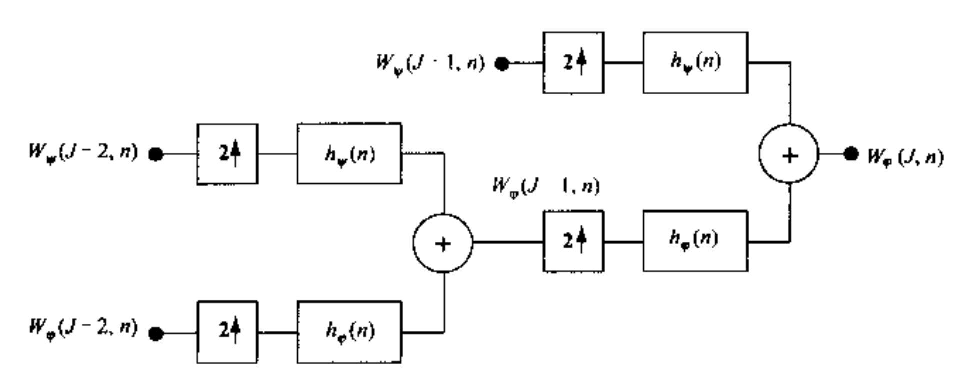


图 7.19 一个二阶或两尺度 FWT^{-1} 的综合滤波器族

例 7.11　计算一维快速小波反变换

快速小波反变换的计算与其正变换的对应部分对称。图 7.20 说明了例 7.10 中使用的序列的处理过程。首先对 0 级近似值和细节系数进行内插以分别产生 $\{1,0\}$ 和 $\{4,0\}$。与滤波器 $g_0(n)=h_\varphi(n)=\{1/\sqrt{2},1/\sqrt{2}\}$ 和 $g_1(n)=h_\psi(n)=\{1/\sqrt{2},-1/\sqrt{2}\}$ 进行卷积得到 $\{1/\sqrt{2},1/\sqrt{2},0\}$ 和 $\{4/\sqrt{2},-4/\sqrt{2},0\}$，相加得到 $W_\varphi(1,n)=\{5/\sqrt{2},-3/\sqrt{2}\}$。这样，图 7.20 中的一级近似值被重建，它与图 7.17 中的近似值计算结果相匹配。用这个方法继续，在第二个综合滤波器族的右端形成 $f(n)$。

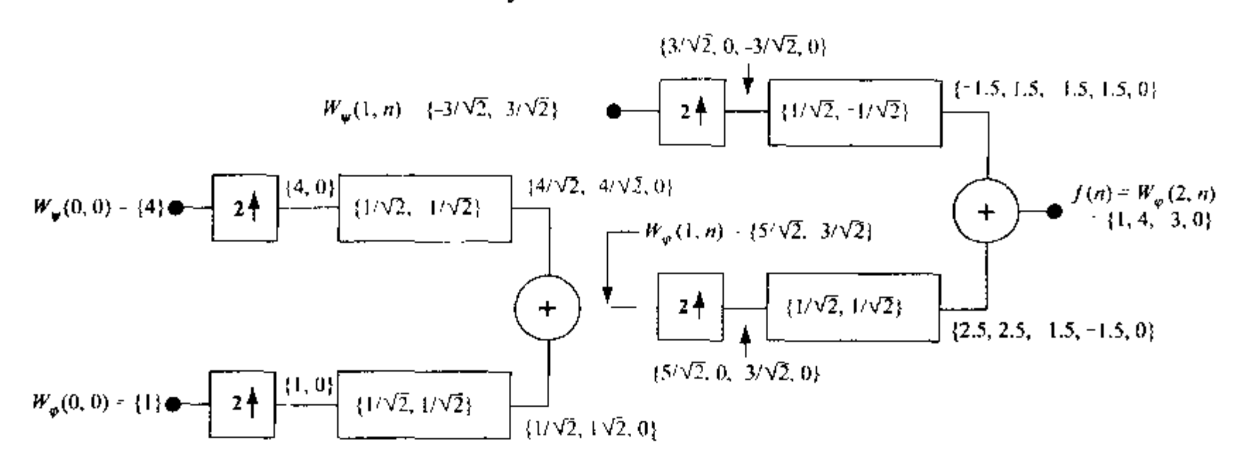


图 7.20　用哈尔尺度和小波向量计算序列 $\{1,4,-1.5\sqrt{2},-1.5\sqrt{2}\}$ 的两尺度快速小波反变换

下面，通过比较 FWT 和 FFT 之间的区别来总结对快速小波变换的讨论。首先是它们数学上的复杂性。包括在长度为 $M=2^J$ 的序列的 FWT 计算中的数学运算次数是 $O(M)$ 阶。也就是说，浮点乘法和加法（使用滤波器族）的次数与序列的长度存在线性关系。这与 FFT 算法比较是有利的，FFT 需要 $O(M\log M)$ 阶。

第二个区别关系到变换的基函数。傅里叶的基函数（即正弦函数）保证了 FFT 的存在，而 FWT 的存在取决于使用的小波的尺度函数是否可用，以及尺度函数和相应的小波的正交性（或双正交性）。因此，式(7.3.12)的墨西哥草帽小波不能用于计算 FWT，因为它没有相应的尺度函数。换句话说，不能为墨西哥草帽小波建立一个像图 7.15 中那样的滤波器族；它不满足 FWT 方法的基本假设。

最后，注意到，虽然在表达函数时时间和频率通常被作为不同的域来处理，但它们之间存在着不可分割的关系。如果试图同时在时域和频域内对函数进行分析，就会遇到如下问题：如果想要关于时域的有价值信息，就要忍受频域的含糊，反之亦然。这是海森伯(Heisenberg)测不准原理在信息处理中的应用。为了用图示说明该原理，函数表达中用到的每个基函数都被概要地看成是时间-频率平面的一个块。该块也称为海森伯单元或海森伯盒，显示了基函数能量的集中区域。块不重叠是正交基函数的特点。

图 7.21 分别显示了时间-频率块：(a)德尔塔冲激函数（传统的时域）基；(b)正弦函数(FFT)基；(c)FWT 基。注意，标准时域基给出了事件发生的时刻，但没有频率信息。而正弦基给出了时间发生的频率但没有时间分辨率。FWT 的时间和频率分辨率是变化的，但各块的面积是相同的。也就是说，每个块都代表了时间-频率平面的一个相等部分。在低频部分，块较短（即，有较好的频域分辨率或较少的频率含糊）但较宽（对应较差的时间分辨率或较多的时间

含糊)。在高频部分,块宽度较小(时间分辨率提高)且块高度变大(表示频率分辨率下降)。FFT 和 FWT 之间的这种基本差别在本章的介绍中就有说明,在频率随时间变化而变化的非稳态函数分析中很重要。

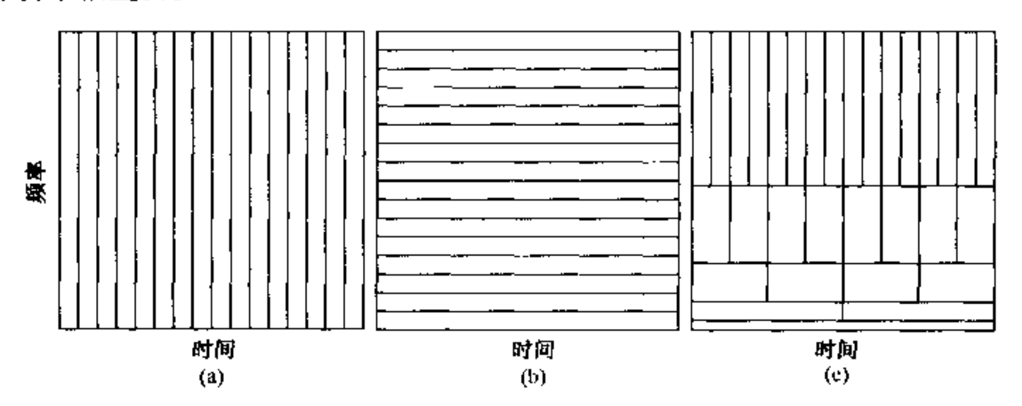


图 7.21 时间-频率块。(a) 取样数据,(b) FFT 基函数,(c) FWT 基函数

7.5 二维小波变换

在前节的一维变换很容易像图像那样扩展到二维函数。在二维情况下,需要一个二维尺度函数 $\varphi(x,y)$和三个二维小波 $\psi^H(x,y)$,$\psi^V(x,y)$和 $\psi^D(x,y)$。每一个都是一维尺度函数 φ 和 ψ 相应的小波函数的乘积。排除产生一维结果的乘积,如 $\varphi(x)\psi(x)$,4 个留下的乘积产生可分离的尺度函数:

$$\varphi(x,y) = \varphi(x)\varphi(y) \tag{7.5.1}$$

和可分离的"方向敏感的"小波:

$$\psi^H(x,y) = \psi(x)\varphi(y) \tag{7.5.2}$$

$$\psi^V(x,y) = \varphi(x)\psi(y) \tag{7.5.3}$$

$$\psi^D(x,y) = \psi(x)\psi(y) \tag{7.5.4}$$

这些小波度量函数会有变化——沿着不同方向的图像强度或灰度的变化:ψ^H 度量沿着列的变化(例如,水平边缘),ψ^V 响应沿着行的变化(例如,垂直边缘),ψ^D 对应于对角线方向的变化。方向敏感是式(7.5.2)到式(7.5.4)所引起的可分离性的自然结果;本节讨论的二维变换的计算复杂性不会增加。

给定可分离的二维尺度和小波函数,一维 DWT 到二维的扩展很简单。首先定义一个尺度和平移基函数:

$$\varphi_{j,m,n}(x,y) = 2^{j/2}\varphi(2^jx - m, 2^jy - n), \tag{7.5.5}$$

$$\psi^i_{j,m,n}(x,y) = 2^{j/2}\psi^i(2^jx - m, 2^jy - n), i = \{H,V,D\} \tag{7.5.6}$$

这里,上标 i 指出式(7.5.2)到式(7.5.4)的方向小波。不像指数,上标 i 代表了值 H,V 和 D。则尺寸为$M \times N$ 的函数$f(x,y)$的离散小波变换是:

$$W_{\varphi}(j_0,m,n)=\frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\varphi_{j_0,m,n}(x,y) \tag{7.5.7}$$

$$W_{\psi}^{i}(j,m,n)=\frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\psi_{j,m,n}^{i}(x,y)\quad i=\{H,V,D\} \tag{7.5.8}$$

如同一维情况，j_0 是任意的开始尺度，$W_{\varphi}(j_0,m,n)$系数定义了在尺度 j_0 的 $f(x,y)$的近似。$W_{\psi}^{i}(j,m,n)$系数对于 $j\geqslant j_0$ 附加了水平、垂直和对角方向的细节。通常令 $j_0=0$ 并且选择 $N=M=2^J$，$j=0,1,2,\cdots,J-1$ 和 $m,n=0,1,2,\cdots,2^j-1$。给出式(7.5.7)和式(7.5.8)的 W_{φ} 和 W_{ψ}^{i}，$f(x,y)$可通过离散反小波变换得到。

$$f(x,y)=\frac{1}{\sqrt{MN}}\sum_{m}\sum_{n}W_{\varphi}(j_0,m,n)\varphi_{j_0,m,n}(x,y)+\frac{1}{\sqrt{MN}}\sum_{i=H,V,D}\sum_{j=j_0}^{\infty}\sum_{m}\sum_{n}W_{\psi}^{i}(j,m,n)\psi_{j,m,n}^{i}(x,y) \tag{7.5.9}$$

类似一维离散小波变换，二维 DWT 可以用数字滤波器和抽样来实现。用可分离的二维尺度和小波函数，我们简单地先取 $f(x,y)$的行的一维 FWT，紧接着用结果列的一维 FWT。图 7.22(a)以方框图的形式显示了这一过程。注意，类似图 7.15 的一维对应部分，二维 FWT“滤波器”尺度 $j+1$ 的近似系数建立了尺度 j 的近似和细节系数。然而，在二维情况下，得到三组细节系数——水平、垂直和对角线细节。

图 7.22(a)的单尺度滤波器族可以用“迭代”(用将近似输出连接到另一个滤波器族的输入的方法)在尺度 $j=J-1,J-2,\cdots,J-P$ 中产生 P 尺度变换。如在一维情况中的，图像 $f(x,y)$被用于 $W_{\varphi}(J,m,n)$的输入。用 $h_{\varphi}(-n)$和 $h_{\psi}(-n)$卷积并对它的列抽样，得到两个子图像，它们的水平分辨率以 2 为因子下降。高通或细节分量描述了图像垂直方向的高频信息，低通近似分量包含它的低频垂直信息。然后，两个子图像以列的方式被滤波并抽样得到 4 个四分之一大小的子图像——W_{φ}，W_{ψ}^{H}，W_{ψ}^{V} 和 W_{ψ}^{D}。示于图 7.22(b)中间的这些子图像是 $f(x,y)$和式(7.5.1)到式(7.5.4)中二维尺度与小波函数的内积，紧接着在每维进行抽样。滤波处理的两次迭代在图 7.22(b)的最右侧产生两尺度分解。

图 7.22(c)显示了上边描述过的反向处理的综合滤波器族。正如所期望的那样，重建算法与一维情况下是相似的。在每一次迭代，四尺度 j 的近似值和细节子图像用两个一维滤波器内插和卷积——一个在子图像的列操作，另一个在行上操作。附加结果是尺度 $j+1$ 的近似值，并且迭代处理一直进行到原始图像被重建。

例 7.12　计算二维快速小波变换

考虑示于图 7.23 的二维 FWT。这里看到图 7.23(a)中由计算机产生的 128×128 基于滤波器分解的一个序列。为了产生这一结果，使用了图7.22(a)的二维滤波器族和图 7.24(a)和(b)的对称小波(symlet)滤波器。图 7.23(b)，(c)和(d)是分解结果。在第一次分解中[图 7.23(b)]，原始图像——在背景上的一系列正弦脉冲——是滤波器族的输入；在所有子序列分解中，近似图像 W_{φ}——前边分解的左上角的子图像——是输入。每一次分解产生 4 个四分之一大小的输出图像，它们如图 7.22(b)所示安排并代替导出它们的输入。注意，基于小波的子图像在每一个尺度上的方向特性，W_{ψ}^{H}，W_{ψ}^{V} 和 W_{ψ}^{D}。

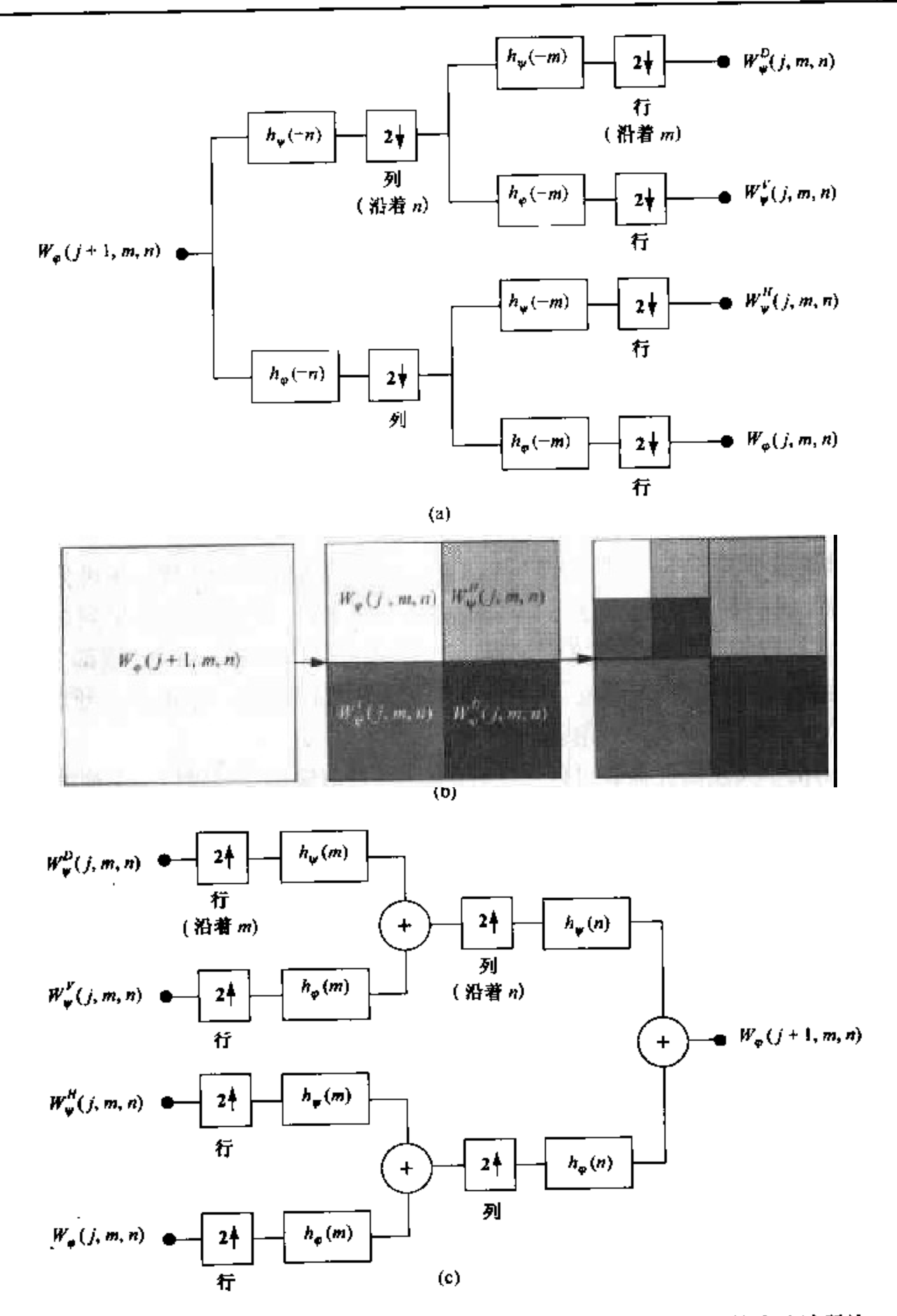


图 7.22 二维快速小波变换。(a) 分析滤波器族,(b)分解结果,(c) 综合滤波器族

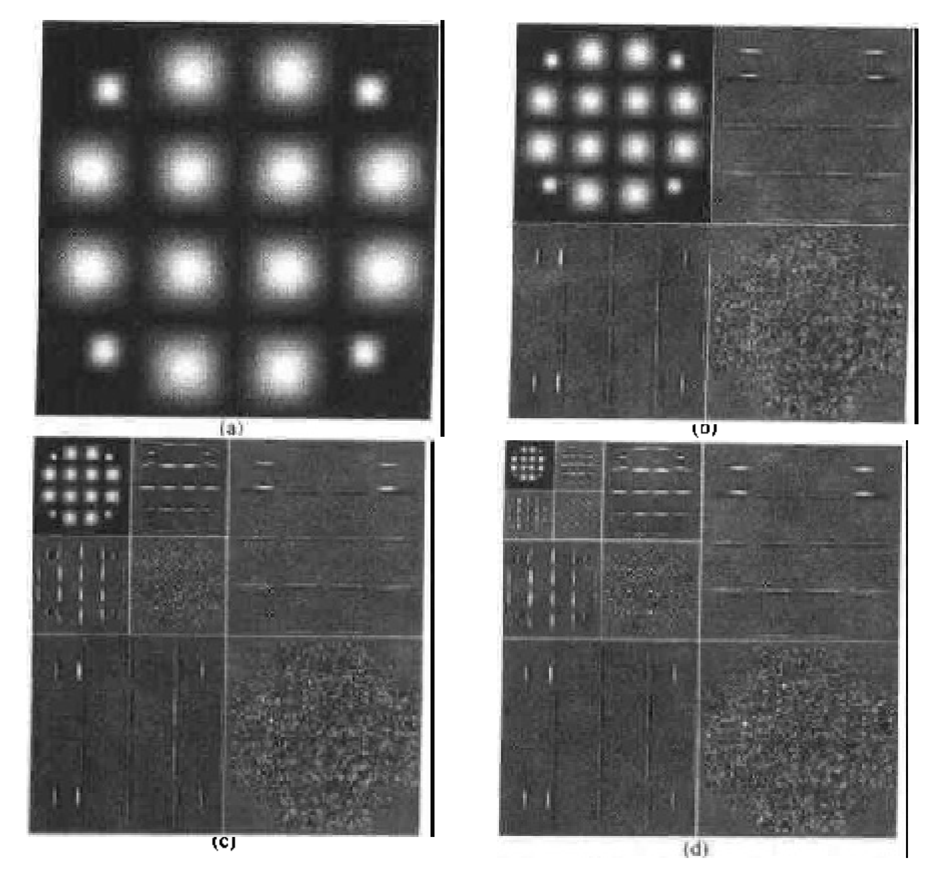


图 7.23　3 尺度 FWT

用于前面例子的分解滤波器是众所周知的称为“对称小波”的小波族，symlet 是对称小波的英文简称。虽然它们不完全对称，但它们在给定的紧支撑下(Daubechies[1992])被设计为有最小不对称性和最高消失矩数①。图 7.24(e)和(f)显示了 4 阶一维对称小波(即：小波和尺度函数)。图 7.24(a)到(d)显示了相应的分解和重建滤波器。低通重建滤波器 $g_0(n)=h_\varphi(n)$ 的系数对于 $0\leqslant n\leqslant 7$ 是 0.032 2, -0.012 6, -0.099 2, 0.297 9, 0.803 7, 0.497 6, -0.029 6 和 -0.075 8。其余的正交滤波器的系数可用式(7.1.23)得到。作为小波 $\psi^H(x,y)$ 的低分辨率图解，图 7.24(g)说明了一维尺度和小波函数怎样结合以形成可分离的二维小波。

我们用两个示范图像处理中小波用途的例子结束本节。如在傅里叶域那样，基本方法是：

1. 计算一幅图像的二维小波变换。
2. 修改变换。
3. 计算反变换。

因为 DWT 尺度和小波函数表现为低通和高通滤波器特性，大多数基于傅里叶滤波的技术与小波部分是等价的。

① 小波 $\psi(x)$ 的第 k 阶矩是 $m(k)=\int x^k\psi(x)\mathrm{d}x$。0 阶矩影响小波函数和尺度函数的平滑性及以多项式表示它们的能力。一个 N 阶对称小波有 N 个消失矩。

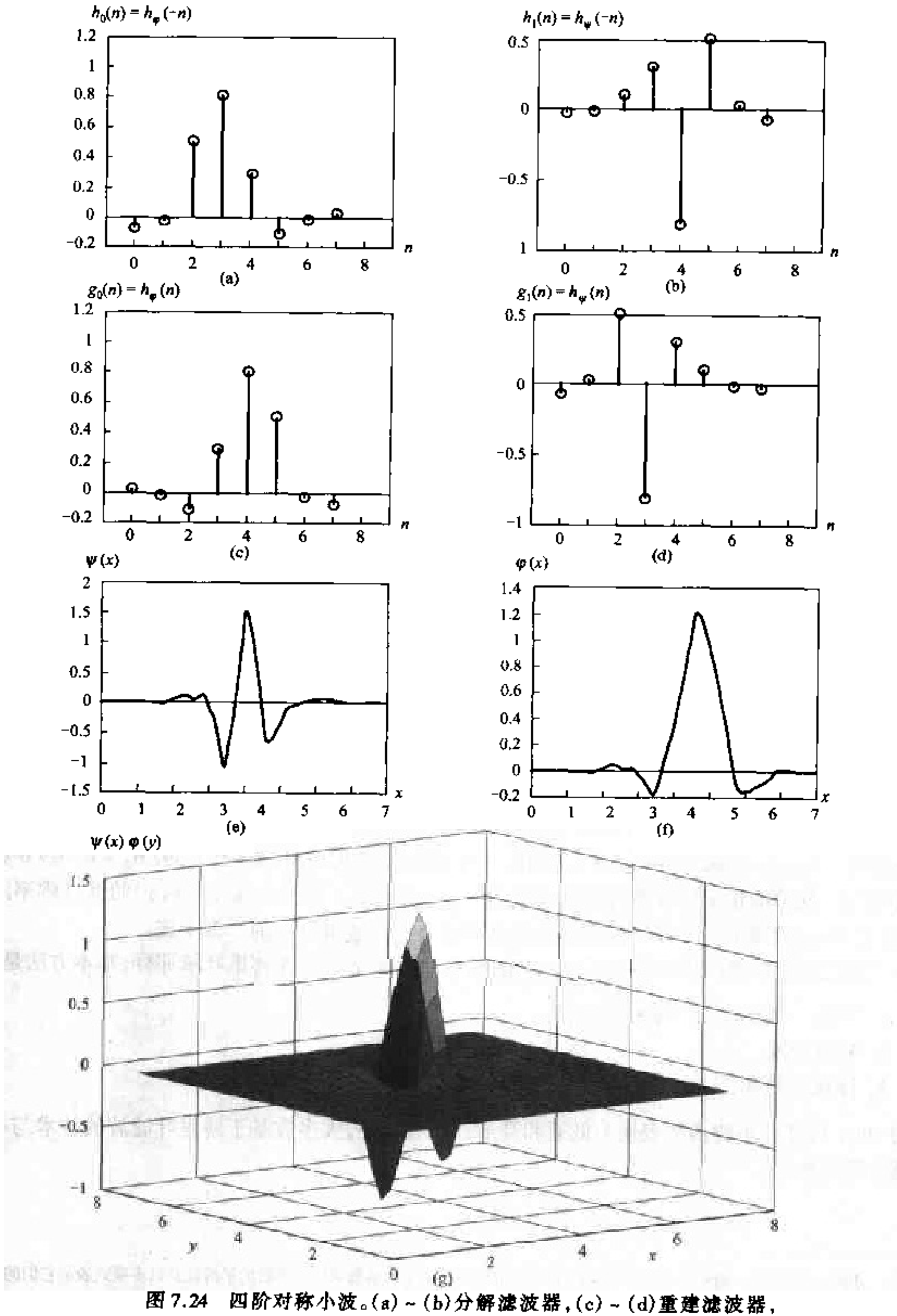


图 7.24 四阶对称小波。(a)~(b)分解滤波器,(c)~(d)重建滤波器,(e)一维小波,(f)一维尺度函数,(g)三个二维小波之一,$\psi^H(x,y)$

例7.13 基于小波的边缘提取

图7.25提供了一个简单的说明。在图7.25(a)中,示于图7.23(c)的离散小波变换的最低尺度近似分量已由置其值为0而消除了。当用这些改进的变换系数计算反变换时,如图7.25(b)所示,效果是强调和突出了图像的边缘。这就可以相当深入地了解原始图像的边缘位置,尽管它们是正弦平移软相关的。参见图7.25(c)和(d),通过水平细节置零,可以孤立垂直边缘。

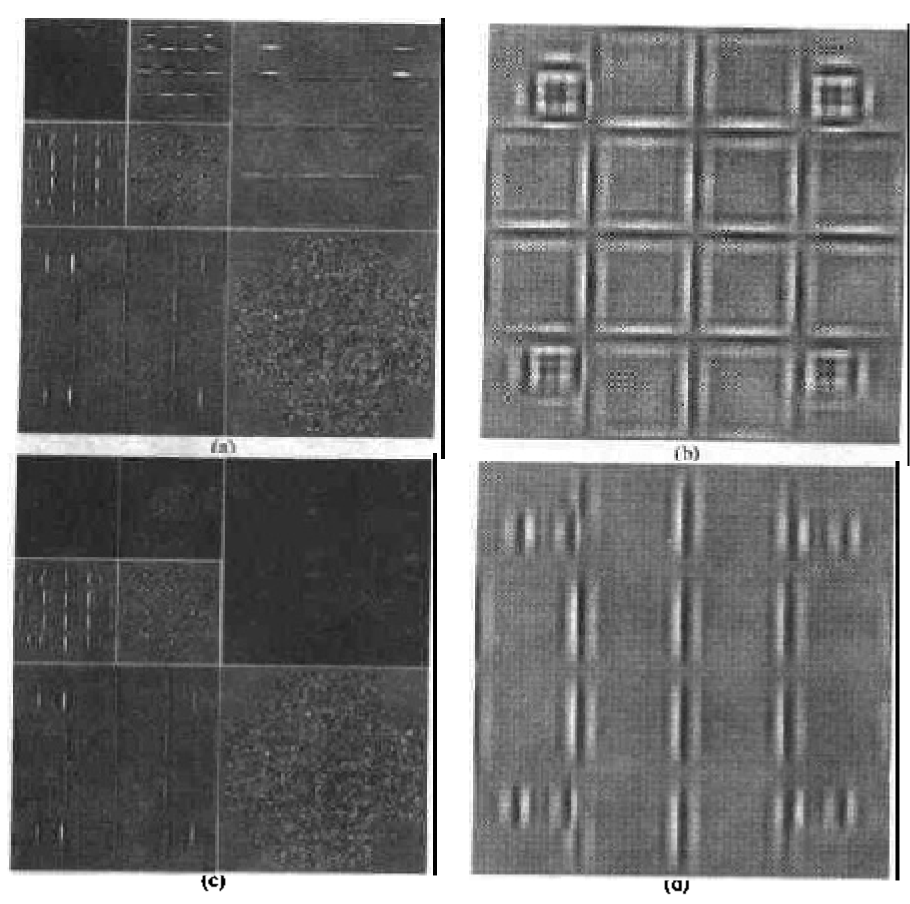


图7.25 改进的边缘检测DWT。(a)和(c)选择的删去系数的两尺度分解,(b)和(d)相应的重建

例7.14 基于小波的噪声去除

作为第二个例子,人脑的核磁共振(MRI)图像显示在图7.26(a)中。正如在背景中所看到的,图像已经被一个加性的或乘性的白噪声干扰了。通常基于小波的对图像去噪声(即,消除噪声部分)的过程如下所示:

1. 为分解,选择一个小波(如,哈尔对称小波)和级别数或尺度 P。然后,计算噪声图像的FWT。
2. 门限化细节系数。即,从尺度 $J-1$ 到 $J-P$ 选择和应用一个门限处理细节系数。这可以由硬门限实现,即元素绝对值比门限值低则置零,或由软门限实现,即元素绝对值比门限值低则置为零,并且标定非零的系数接近零。软门限去除了硬门限所固有的不连续性(在门限处)。
3. 基于原始的近似系数,在 $J-P$ 级执行小波重建,并对 $J-1$ 到 $J-P$ 级改进细节系数。

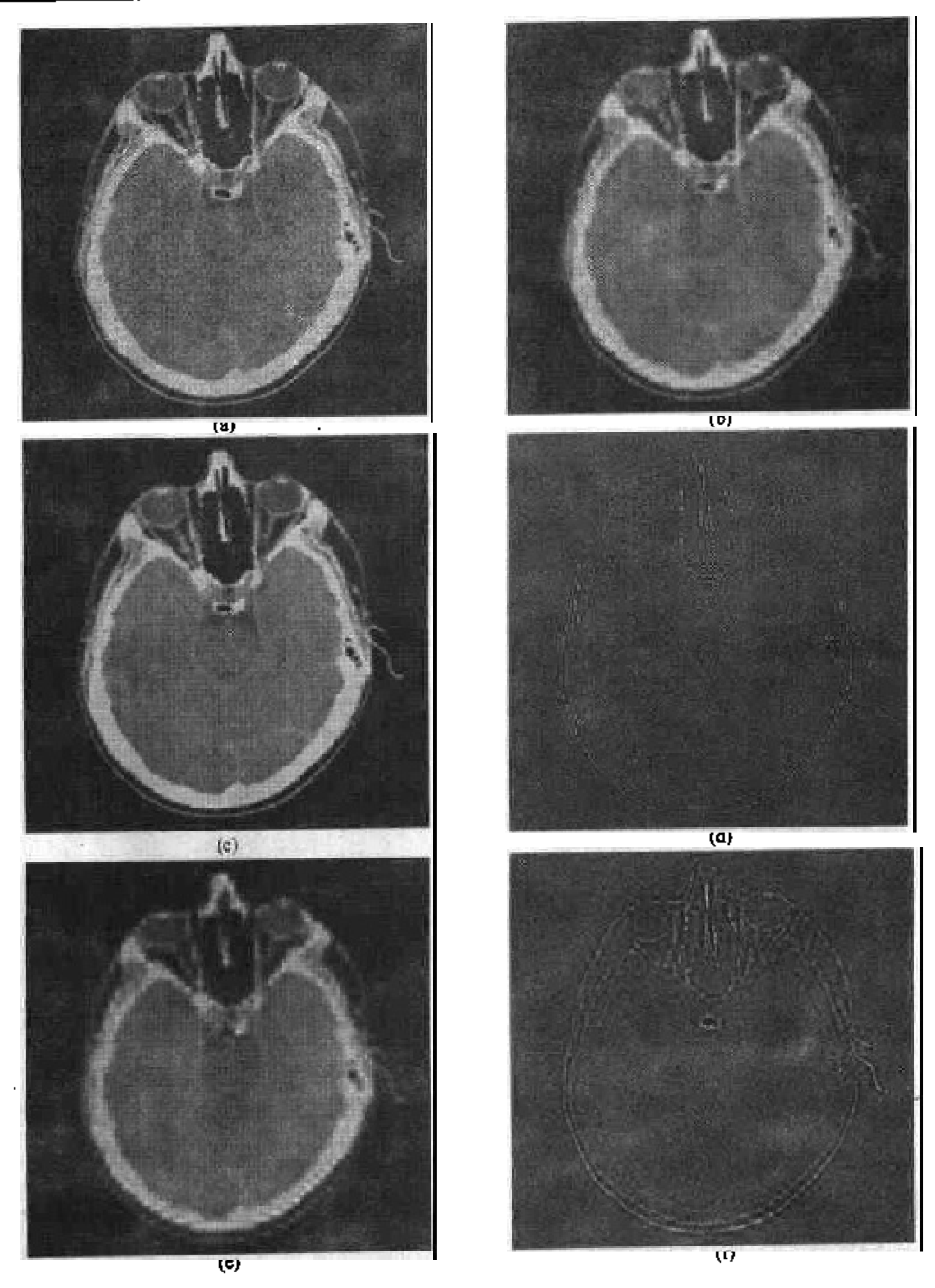


图 7.26 对噪声去除改进 DWT。(a)人体头部的带噪声的 MRI 图像,(b),(c)和(e)门限化细节系数后的各种重建,(d)和(f)在(c)和(e)重建时移去的信息(原图像由 Vanderbuilt 大学的医学中心提供)

图 7.26(b)显示了用第四阶对称小波、2 尺度($P=2$)、全局门限 94.909 3 执行这一操作的结果。注意,噪声的减少和相应的图像边缘质量的损失。这一边缘细节的损失在图 7.26(c)中大大减少了,它是由最高分辨率细节系数(不是低分辨率)置零并重建图像产生的。这里,几乎所有的背景噪声都被去掉了,而边缘只稍微有点干扰。图 7.26(d)显示了处理中丢失的信

息。这个结果是由计算分解图像所有系数(但高分辨率系数为零)的反 FWT 产生的。正如所看到的,它包含了原图像中大多数噪声和某些边缘信息。图 7.26(e)和(f)是包含所有细节系数的相似图像集合。即,图 7.26(e)是在两个分解级别的细节被置零时的 DWT 重建;图 7.26(f)是细节系数(即,仅最低级近似系数)被去除的重建。注意,在图 7.26(f)中边缘信息的显著增加和图 7.26(e)中边缘细节的相应减少。

7.6 小波包

快速小波变换将一个函数分解为一系列与对数相关的频段。即,低频被组成窄频段,高频被组成宽频段。如果沿着图 7.21(c)中时频平面的频率轴看,这立刻就清楚了。它规定了通常所谓的恒定 Q 滤波器。如果想要较大地控制时频平面的一部分(即,高频较相似的频段),FWT 必须由更灵活的分解——称为小波包(Coifman 和 Wickerhauser[1992])产生。这个产生过程的代价是 FWT 计算复杂性的增加,从 $O(M)$到 $O(M\log M)$。

再一次考虑图 7.16(a)的两尺度滤波器族,但是,图示分解是一个二叉树。图 7.27(a)详细叙述了树的结构并将合适的 FWT 尺度和小波系数[从图 7.16(a)]连接到它的节点。根节点被赋予最高的尺度近似系数,它是函数自身的取样,但叶子继承变换的近似和细节系数的输出。单独的中间节点 $W_\varphi(J-1,n)$是滤波器族近似,它最终被滤波为叶子节点。注意,每一个节点的系数都是线性扩展的权重,它产生一个根节点 $f(n)$的带限的"片"。因为任何这样的"片"都是一个已知的尺度或小波子空间的元素(见 7.2.2 节和 7.2.3 节),所以可以用相应的子空间替换以产生图 7.27(a)的系数。结果是图 7.27(b)的子空间分析树。虽然采用了可变 W 表示系数和子空间,但两个数值按照它们的下标形式是可区分的。

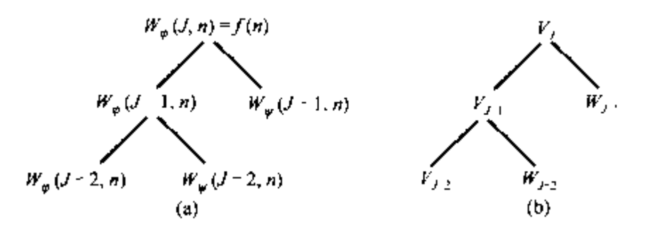


图 7.27　对于图 7.16 的两尺度 FWT 分析族的一个系数(a)和分析(b)树

这些概念进一步说明于图 7.28,其中,描绘了一个三尺度 FWT 分析族、分析树和相应的频谱。不像图 7.16(a),图 7.28(a)的方框图标明类似图 7.28(b)中的分析树,以及图 7.28(c)的频谱。当左上边的滤波器和子取样器的输出是准确的 $W_\psi(J-1,n)$时,它被标为 W_{J-1}——函数的子空间,它是由 $W_\psi(J-1,n)$变换系数产生的。这个子空间对应于相联系的分析树左上边的叶子,以及对应频谱的最右边或最宽带宽段。

分析树提供了表示多尺度小波变换的紧凑和有益的方法。它们比相对应的滤波器和基于子取样的方框图容易画并占有较少的空间,使它相对容易定位有效的分解。例如,图 7.28(b)的三尺度分析树提供了下列三种展开选择:

$$V_J = V_{J-1} \oplus W_{J-1} \tag{7.6.1}$$

$$V_J = V_{J-2} \oplus W_{J-2} \oplus W_{J-1} \tag{7.6.2}$$

$$V_J = V_{J-3} \oplus W_{J-3} \oplus W_{J-2} \oplus W_{J-1} \qquad (7.6.3)$$

它们对应7.4节的一、二、三尺度FWT的分解，并且可以通过7.2.3节的式(7.2.27)，对 $P = \{1,2,3\}$ 令 $j_0 = J - P$ 得到。通常，P 尺度的FWT分析树支持 P 惟一的分解。

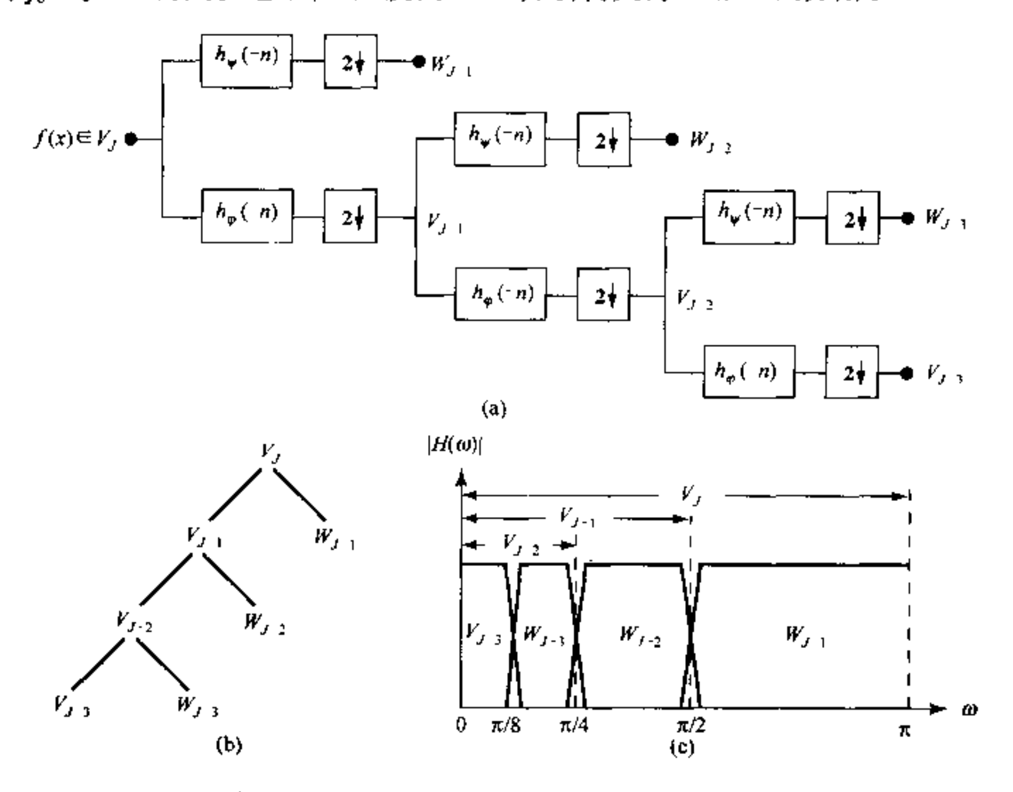


图7.28　三尺度FWT滤波器族。(a)方框图，(b) 分解空间树，(c)谱分离特性

分析树还是表示小波包的有效机理，它们比在细节是迭代滤波器的传统小波变换更简单。这样，图7.28(b)的三尺度FWT分析树变成图7.29的三尺度小波包树。注意，引入的附加下标。双下标节点的第一个下标指出FWT父节点的尺度。第二个是从父节点到该点通路的 A 和 D 编码的可变长度串。A 表示近似值滤波，而 D 表示细节滤波。例如，子空间 $W_{J-1,DA}$ 是由尺度 $J-1$ 的FWT系数(如图7.29中的父节点 W_{J-1})通过一个附加细节滤波器(生成 $W_{J-1,D}$)"进行滤波"，而后紧跟着一个近似值滤波器(给出 $W_{J-1,DA}$)得到的。图7.30(a)和(b)是图7.29中分析树的滤波器组和谱线分离特征。注意平均分布的频带，这是完全小波包分解的特征。

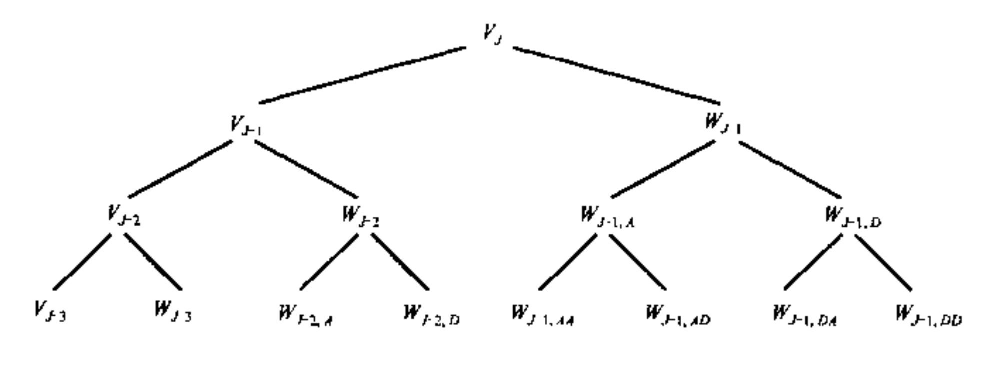


图7.29　一个三尺度小波包分析树

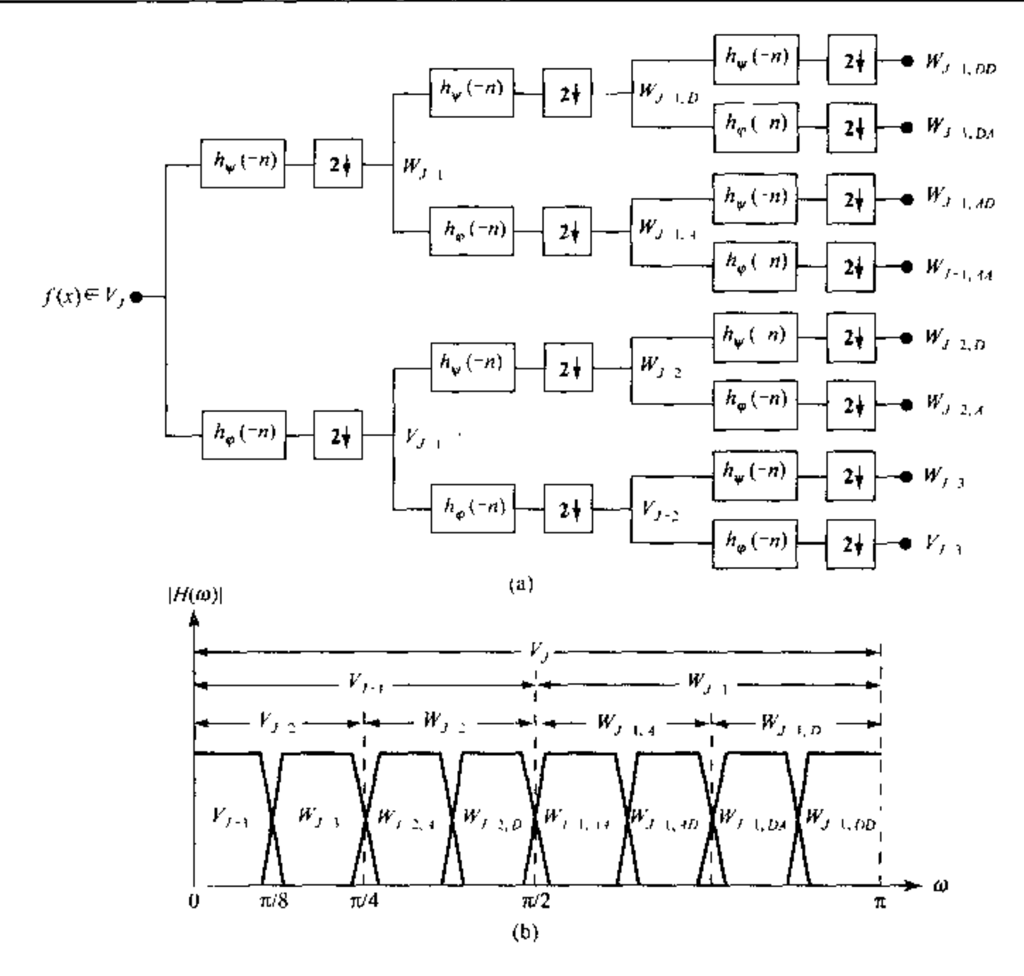


图 7.30 (a)滤波器组和(b)一个三尺度全小波包分析树的频谱分离特征

图 7.29 中的三尺度小波包树几乎是三尺度 FWT 树的有效分解(和相关的时间-频率片)数目的三倍。回顾在正常的 FWT 中,我们进行分离、滤波,并且单独对低通波段进行抽样。这将在频率波段间生成一个固定的对数函数关系。因此,当图 7.28(a)中的三尺度 FWT 分析树提供三种可能的分解的时候——见式(7.6.1)到式(7.6.3)——图 7.29 的小波包树支持 26 种不同的分解。例如,V_J[及函数 $f(n)$]可以进行如下扩展:

$$V_J = V_{J-3} \oplus W_{J-3} \oplus W_{J-2,A} \oplus W_{J-2,D} \oplus W_{J-1,AA} \oplus W_{J-1,AD} \oplus W_{J-1,DA} \oplus W_{J-1,DD} \quad (7.6.4)$$

它们的频谱如图 7.30(b)中所示,或者:

$$V_J = V_{J-1} \oplus W_{J-1,D} \oplus W_{J-1,AA} \oplus W_{J-1,AD} \quad (7.6.5)$$

它们的频谱在图 7.31 中进行了描述。注意,最后的频谱和图 7.30(b)的完全包频谱或图7.28(c)的三尺度 FWT 频谱之间的区别。一般来说,P 尺度、一维小波包变换(和 $P+1$ 层分析树相关)支持惟一的分解,这里 $D(1)=1$。

$$D(P+1) = [D(P)]^2 + 1 \quad (7.6.6)$$

随着这样大量的有效扩展,基于包的变换改进了对被分解函数的频谱分割的控制。当然,这种控制的代价就是计算复杂性的增加[将图 7.28(a)中的滤波器组与图 7.30(a)的进行比较]。

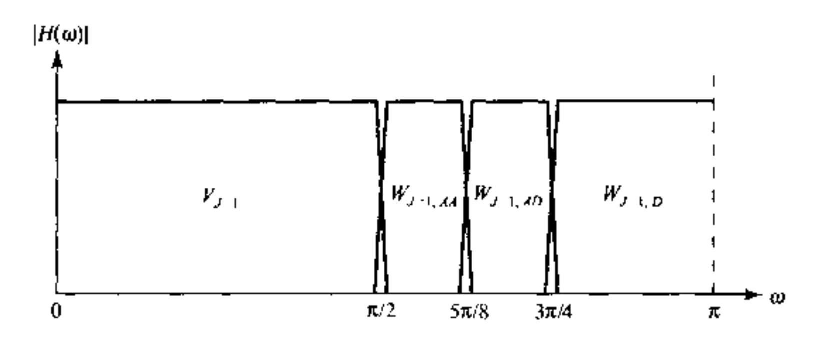


图 7.31 式(7.6.5)中分解的频谱

现在考虑一下图 7.22(a)中的二维、四波段滤波器组。正如在 7.5 节中提到的那样,这个滤波器组将近似值 $W_\varphi(j+1,m,n)$分解成 $W_\varphi(j,m,n)$, $W_\psi^H(j,m,n)$, $W_\psi^V(j,m,n)$和 $W_\psi^D(j,m,n)$输出。在一维条件下,这个过程可以对 $j=J-1,J-2,\cdots,J-P$,同 $W_\varphi(J,m,n)=f(m,n)$一起反复迭代生成 P 尺度变换。由第一次迭代[即,在图 7.22(a)中使用$j+1=J$]得到的频谱示于图 7.32(a)。注意,它将频率平面分割成 4 个相等的区域。位于平面中心的低频四分之一波段与变换系数 $W_\varphi(J-1,m,n)$及尺度空间 V_{J-1}相一致(这种命名法适用于一维的情况)。然而,为适应输入的二维特性,就要有 3 个(而不是一个)小波子空间。它们分别用 W_{J-1}^H, W_{J-1}^V以及 W_{J-1}^D表示,对应的系数分别为 $W_\psi^H(J-1,m,n)$, $W_\psi^V(J-1,m,n)$和 $W_\psi^D(J-1,m,n)$。图 7.32(b)显示了得到的四波段、单尺度的 4 个 FWT 分析树。请注意上标,正是它们将小波子空间号和它们对应的二维变换系数联系在一起的。

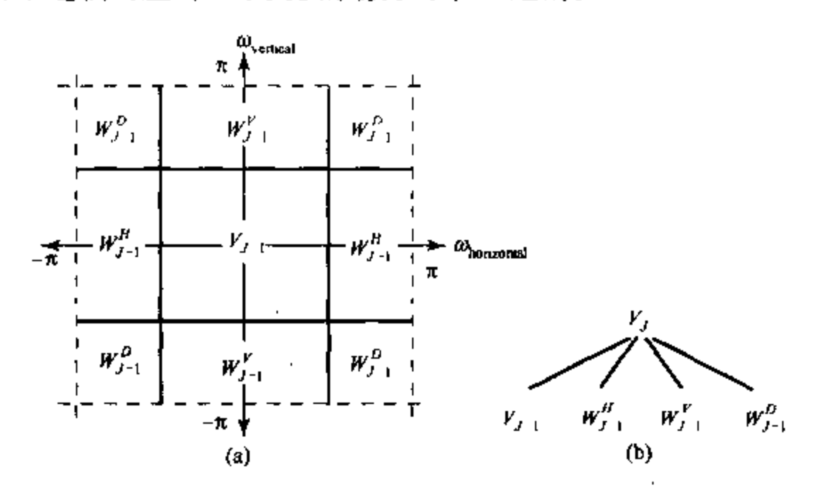


图 7.32 一个二维 FWT 的第一次分解。(a)频谱,(b)子空间分析树

图 7.33 显示了一个三尺度、二维小波包分析树的一部分。像图 7.29 中它对应的一维部分一样,传统 FWT 细节节点下一代的每一个节点的第一个下标都是其父细节节点的尺度。第二个下标——一个 A 和 D 的可变长度字串——对从父节点到当前节点的路径进行了编码。例如,标记为 $W_{J-1,VD}^H$的节点是通过首先用一个附加的细节/近似值滤波器(生成 $W_{J-1,V}^H$),然后使用一个细节/细节滤波器(得到 $W_{J-1,VD}^H$)对尺度 $J-1$ 的 FWT 水平细节系数(如图 7.33 中的父节点 W_{J-1}^H)进行"行/列"滤波得到的。一个 P 尺度、二维小波包树支持:

$$D(P+1)=[D(P)]^4+1 \tag{7.6.7}$$

惟一的展开，这里 $D(1)=1$。因此，图 7.33 的三尺度树提供了 83 522 种不同的分解！

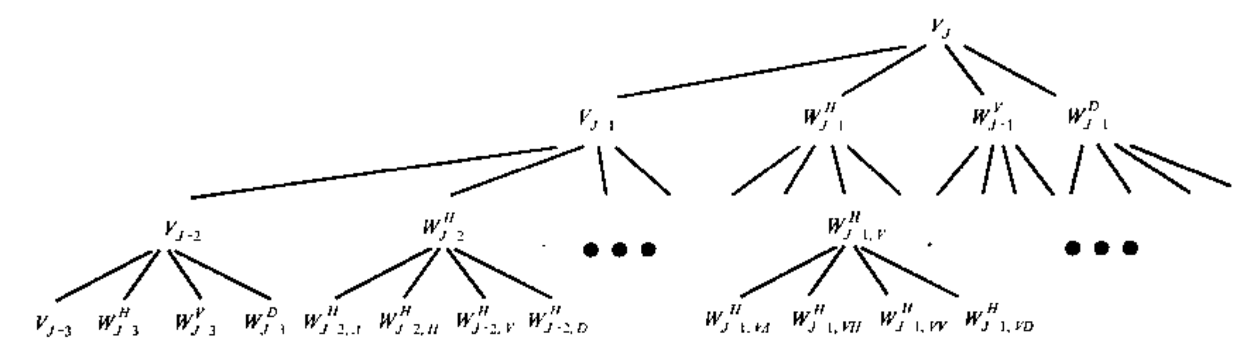


图 7.33　一个三尺度、全小波包分解树。只给出了树的一部分

例 7.15　二维小波包分解

单一的小波包树存在多种分解选择。实际上，可能的分解数目通常非常大，无法实际将每一种分解都列举出来或进行检验。非常需要一种有效的算法来寻找适用于特定应用准则的最佳分解。正如将要看到的，经典的以熵为基础的准则在许多场合下是适用的，并且较为适合应用于二叉和四叉树搜索算法。

考虑图 7.34(a)中的指纹图像的压缩问题。使用三尺度小波包树，有 83 522[见式(7.6.7)]种可能的分解方式，每种分解都可以作为压缩操作的起点。

图 7.34(b)显示了其中的一种分解方式——一种全小波包，分解为 64 片，类似图 7.33 的分析树。注意，树的叶子对应图 7.34(b)中被分解的子图构成的 8×8 阵列的子带。然而，以压缩为目的使这种特定的 64 片分解达到某种程度优化的可能性是相对较低的。在缺乏一种合适的最佳化准则的时候，无法肯定或否定这种分解是否是最优的。

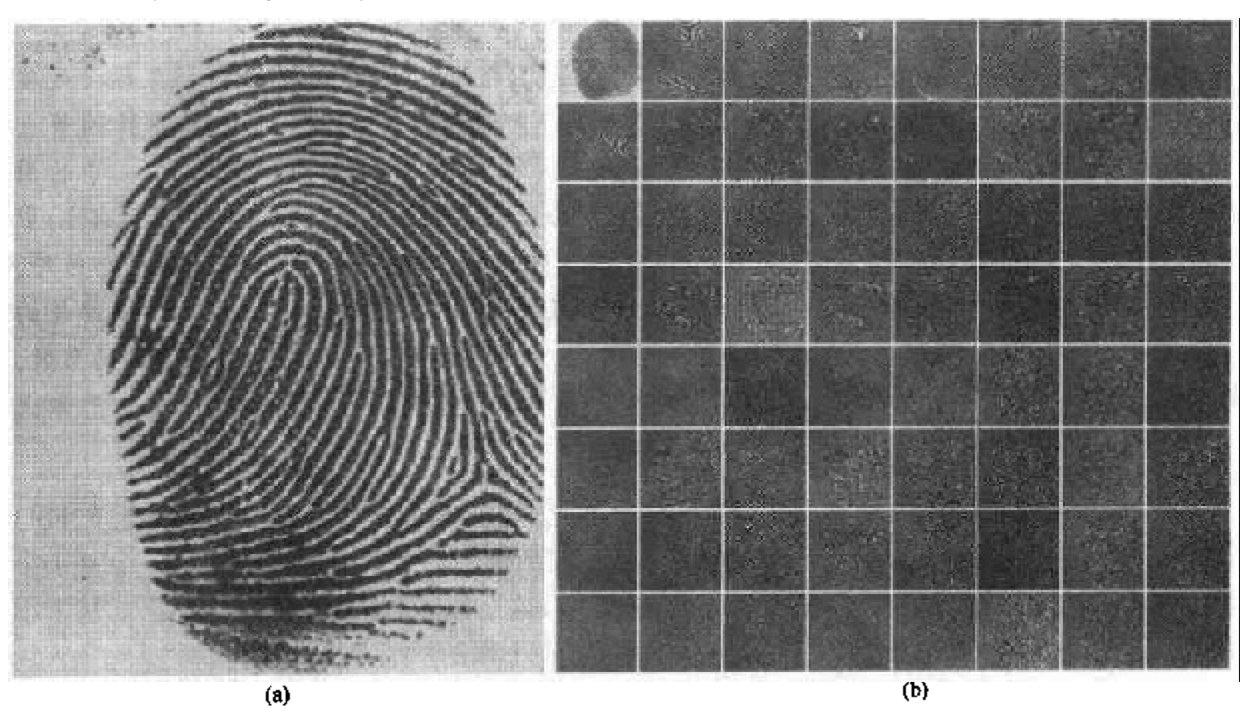


图 7.34　(a)一幅指纹扫描图像，(b)它的三尺度、全小波包分解(原图像由美国国家标准与技术协会提供)

对图 7.34(a)图像的压缩选择一种分解方式的较为合理的准则是附加代价函数：

$$E(f) = \sum_{m,n} |f(m,n)| \tag{7.6.8}$$

这个函数测量了二维函数 f 的熵或信息容量。熵接近 0 表明函数带有极少的信息量或不携带任何信息。例如，函数 $f(m, n) = 0$ 对所有的 m 和 n 的熵均为 0。这个函数就没有提供任何信息。另一方面，熵值很高，表明函数具有很多非零值。由于大多数基于变换的压缩方案都是通过舍入或设置门限将很小的系数变成零值的方式工作，因此，从压缩操作的观点来看，将接近零值的数目最大化的代价函数对于最好的分解应该是合理的准则。

刚才提到的代价函数不仅具有计算上的简单性，而且易于应用在树的优化程序中。最优化算法必须使用这个函数使分解树的叶子节点所付的"代价"最小。熵最小的节点应该是最受欢迎的，因为这些节点具有更多的接近零的数值，从而导致更大的压缩。因为式(7.6.8)这个代价函数只是一个局部度量，它只用到考虑之中的节点的有效信息，所以一个用于得到最小熵的有效算法很容易构造。

对于分析树的每个节点，要从根节点开始逐层构造树，直到到达叶子节点：

1. 计算节点的熵[用 E_P(作为父节点的熵)表示]和此节点的 4 个子节点的熵——分别表示为 E_A，E_H，E_V 和 E_D。对于二维小波包分解，父节点是一个近似值或细节系数的二维阵列；它的子节点是经过滤波的近似值，是水平、垂直和对角线方向上的细节。
2. 如果子节点的联合熵小于父节点的熵——即 $E_A + E_H + E_V + E_D < E_P$——就将这些子节点包括到分析树中，如果子节点的联合熵大于或等于父节点的熵，就去掉这些子节点，只保留父节点。此节点就是最佳分析树的一个叶子节点。

前述算法可以用于：(1)对小波包树进行修剪，或(2)根据计算最优树的大致框架设计处理程序。在后一种情况中，不必要的兄弟节点——那些在算法的第二步中将去掉的节点后代——就不用计算了。图 7.35 显示出使用式(7.6.8)的代价函数对图 7.34(a)中的图像应用此算法得到的最优化分解。图 7.36 中给出了对应的分析树。注意，图7.34(b)中多个原始全包分解的 64 个子带(图 7.33 中相应分析树的 64 个叶子)已被去除了。另外，图 7.35 中没有被分离(或进一步分解)的子图像比较平滑并且由那些具有中间灰度值的像素构成。由于这幅图的近似值子图已被标定，所以灰度级 128 表示一个值为零的系数，这些子图几乎不包含任何信息。对这些子图的分割应该不会使图像的熵在总体上减少。

前述例子是以实际问题为根据的，它可以通过使用小波得到解决。美国联邦调查局(FBI)近年来拥有一个巨大的指纹数据库，并且已经为指纹图像的数字化及压缩制定了一个基于小波的国家标准(FBI[1993])。由于使用双正交的小波，这个标准达到了15:1的压缩率。基于小波的压缩相对于更为传统的JPEG压缩方法的优点会在下面的章节中得到检验。

在例7.15中使用的分解滤波器，同 FBI 使用的一样都是著名的称为Cohen-Daubechies-Feauveau的双正交小波族(Cohen，Daubechies和Feauveau[1992])中的组成部分。因为这个滤波器族的尺度和小波函数是对称的并且具有相近的长度，所以它们是使用最为广泛的双正交小波。

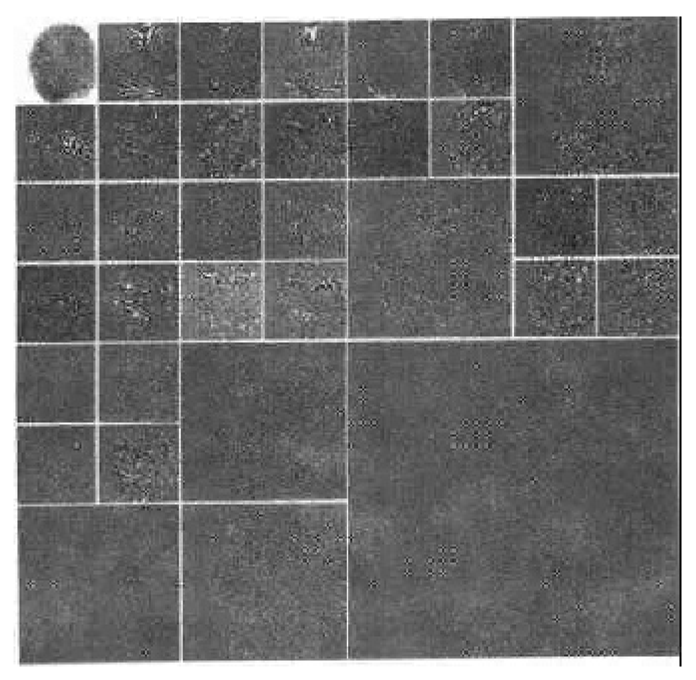


图 7.35　对图 7.34(a)中的指纹进行的一种最佳小波包分解

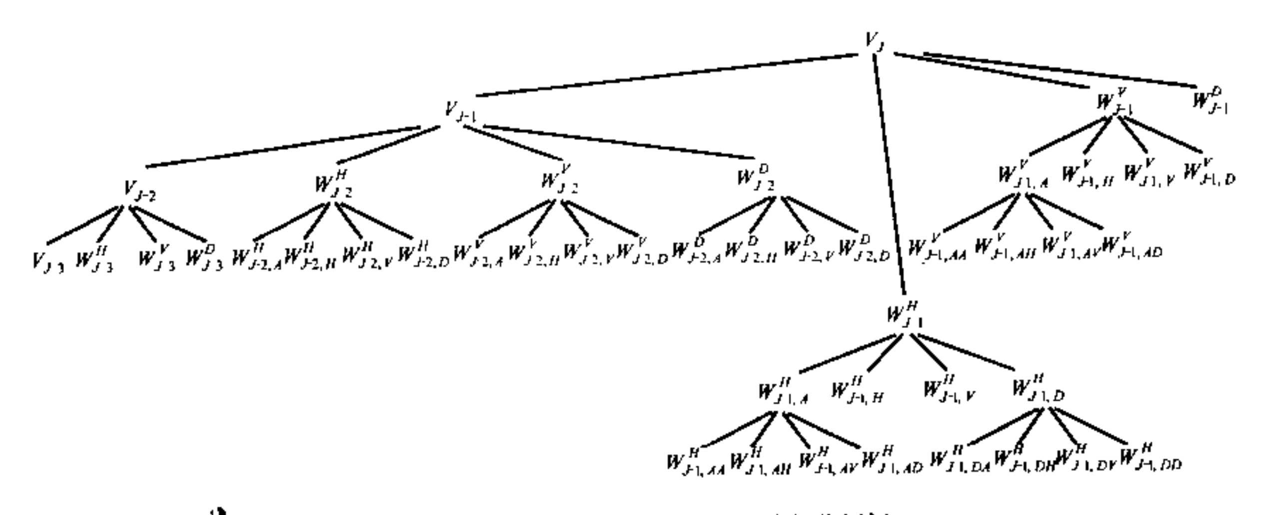


图 7.36　图 7.35 中分解的最佳小波包分析树

图 7.37(e)到(h)显示了双尺度和以双正交这一要素为特性的小波函数。图 7.37(a)到(d)是对应的分解和重构滤波器。当 $0 \leqslant n \leqslant 17$ 的时候，低通分解滤波器 $h_0(n)$ 的系数为 0,0.001 9, −0.001 9, −0.017,0.011 9,0.049 7, −0.077 3, −0.094 1,0.420 8,0.825 9,0.420 8, −0.094 1, −0.077 3,0.049 7,0.011 9, −0.017, −0.001 9 和 0.001 0。当 $0 \leqslant n \leqslant 17$ 的时候，高通分解滤波器 $h_1(n)$ 的系数为 0,0,0,0.014 4, −0.014 5, −0.078 7,0.040 4,0.417 8, −0.758 9,0.417 8, 0.040 4, −0.078 7, −0.014 5,0.014 4,0,0,0 和 0。双正交综合滤波器的对应系数可以使用式(7.1.15)的 $g_0(n) = (-1)^{n+1} h_1(n)$ 和 $g_1(n) = (-1)^n h_0(n)$ 计算得到。就是说，这些系数是分解滤波器交叉调制得到的。注意，填充零是为了使滤波器具有相同的长度。

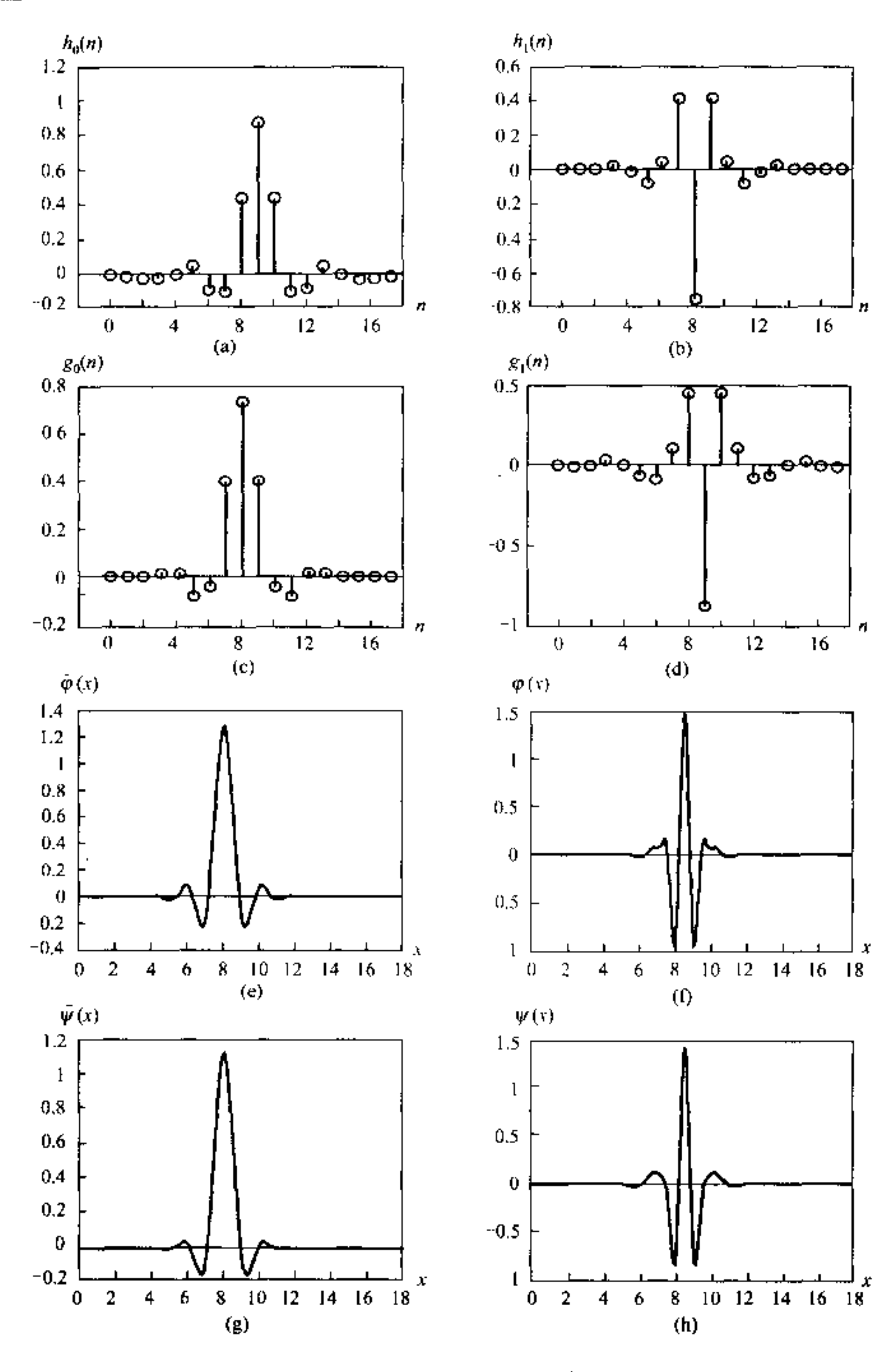


图 7.37 Cohen-Daubechies-Feauveau 双正交小波族的部分小波。(a)和(b)分解滤波器系数，(c)和(d)重构滤波器系数，(e)到(h)双小波及尺度函数

小结

这一章所使用的材料为理解和进一步掌握在图像处理中小波和多分辨率分析所起的作用奠定了牢固的数学基础。小波和小波变换是在大量图像处理课题中快速崛起的较新的图像工具。由于它们与傅里叶变换有相似性,第 4 章中的许多技术同样有基于小波域的部分。成像应用中能够以小波观点来解决的问题包括图像匹配、配准、分割、降噪、重建、增强、压缩、形态滤波和计算机断层摄影术等。由于不可能在一个章节中覆盖所有这些应用领域,所以根据领域的价值只选择了其中的部分主题进行介绍或阐明基本概念,同时为读者在某个领域中进一步研究做好准备。在第 8 章,将使用小波进行图像压缩。

参考资料

关于小波及其应用有许多好的文章。其中一些可以作为我们论述的补充,并且本章核心部分进一步的展开也仰赖于这些文章。在 7.1.2 节中关于子带编码和数字滤波的材料是以Vetterli和Kovacevic[1995]的书为依据的,而 7.2 节和 7.4 节关于多分辨率展开和快速小波变换的部分遵循了Burrus, Gopinath和Guo[1998]著述中的论述。本章中剩下的材料是以文章中引用的参考资料为基础的。实验性的工作和许多实例是使用 MATLAB 小波工具箱做出来的(Misiti,Misiti,Oppenheim和Poggi[1996])。

Hubbard[1998]写的一本书中记述了小波分析的历史。小波的前身同时在不同的领域中得到发展并在Mallat[1987]的文章中得到了统一。这部著述为此领域建立了一个数学框架。从Meyer[1987][1990][1992a,b][1993],Mallat[1987][1989a ~ c][1998],以及Daubechies[1988][1990][1992][1993][1996]的工作中可以追溯小波的发展史。众多关于小波的出版物激发了当前对于小波领域的研究兴趣。Daubechies[1992]所写的书是一部关于小波理论的数学细节方面的经典资料。

在一般的图像处理的文章中都会提及小波在图像处理中的应用,像Castleman[1996],以及许多特定应用领域的书籍,其中还有一些参考学报。例如,在后者中有Rosenfeld[1984],Prasad和Iyengar[1997],还有Topiwala[1998]的著述。最近的一些文章可以作为在一些特定的图像应用方面进一步研究的起点读物,包括Thévenaz和Unser[2000]关于图像配准的文章;Chang 和 Kuo[1993]以及Unser[1995]基于纹理分类的文章;Heijmans和Goutsias[2000]关于形态学小波的文章;Banham等人[1994],Wang,Zhang 和 Pan[1995],还有Banham和Kastaggelos[1996]关于图像复原的文章;Xu 等人[1994]与Chang,Yu和Vetterli[2000]关于图像增强的文章;Delaney和Bresler[1995]以及Westenberg和Roerdink[2000]关于计算机断层摄影术的文章;Lee,Sun和 Chen[1995],Liang 和 Kuo[1999],Wang,Lee和Toraichi[1999],以及 You 和Bhattacharya[2000]关于图像描述和匹配的文章。小波的最重要应用之一是图像压缩——参考例如Antonini等人[1992],Wei 等人[1998]以及Topiwala[1998]的著述。并且已经有不少关于小波的专门刊物,包括IEEE信息论学报[1992]中关于小波变换和多分辨率信号分析的特刊,IEEE信号处理学报[1993]中关于小波和信号处理的特刊,IEEE模式分析和机器智能学报[1989]中关于多分辨率表示的特刊。

尽管本章关注的焦点是小波的基本原理和在图像处理上的应用,但小波本身的构建还是

极有意思的。有兴趣的读者可以参考Battle[1987][1988],Daubechies[1988][1992],Cohen和Daubechies[1992],Meyer[1990],Mallat[1989b],Unser,Aldroubi和Eden[1993],以及Gröchenig和Madych[1992]的著述。这里并没有给出详尽的资料目录,但是这些可以作为进一步阅读的起点读物。也可以参考一些一般的关于子带编码和滤波器组的参考资料,包括Strang和Nguyen[1996]与Vetterli和Kovacevic[1995],以及包含在本章中作为例子给出的参考资料。

习题

7.1 设计一个系统用图 7.2(b)中的编码器对预测残差金字塔进行解码,并画出框图。假设编码器不会产生量化误差。

★**7.2** 对下列图像构建全组合的近似值金字塔和对应的预测残差金字塔:

$$f(x,y)=\begin{bmatrix}1&2&3&4\\5&6&7&8\\9&10&11&12\\13&14&15&16\end{bmatrix}$$

使用相邻 2×2 的块对图 7.2(b)中的近似值滤波器进行平均,并假设忽略插值滤波器。

★**7.3** 给定 $2^J\times2^J$ 的图像,那么一个 $J+1$ 级金字塔是减少还是增加了表示图像所需的数据量?压缩或扩展率是多少?

7.4 证明下列来自表 7.1 的滤波器形成了理想的重构滤波器组:

★(a)正交镜像滤波器(QMF)

(b)正交滤波器

7.5 正交镜像滤波器是双正交的还是正交的,或两者均是?

7.6 计算例7.2的 Daubechies 综合滤波器 $g_0(n)$ 和 $g_1(n)$ 的系数。令 $m=0$,使用式(7.1.22)说明滤波器是正交的。

★**7.7** 画一个二维四波段滤波器组解码器来重构图 7.5 中的输入 $x(m,n)$。

7.8 构造 $N=8$ 的哈尔变换矩阵。

7.9 (a)计算下列 2×2 图像的哈尔变换:

$$\mathbf{F}=\begin{bmatrix}3&-1\\6&2\end{bmatrix}$$

(b)逆哈尔变换是 $\mathbf{F}=\mathbf{H}^{-1}\mathbf{T}\mathbf{H}^{-1}$,这里 $\mathbf{T}$ 是哈尔变换,而 $\mathbf{H}^{-1}$ 表示哈尔变换矩阵 $\mathbf{H}$ 的逆矩阵。寻找哈尔变换矩阵 $\mathbf{H}_2$ 的 $\mathbf{H}_2^{-1}$,并用它计算(a)中结果的逆哈尔变换。

7.10 以下列基本要素计算二元组 $[3,2]^T$ 的扩展系数并写出对应的扩展:

★(a)以二元实数集合 $\mathbf{R}^2$ 为基础的 $\varphi_0=[1/\sqrt{2},1/\sqrt{2}]^T$ 和 $\varphi_1=[1/\sqrt{2},-1/\sqrt{2}]^T$。

(b)以 $\mathbf{R}^2$ 为基础的 $\varphi_0=[1,0]^T,\varphi_1=[1,1]^T$ 和它的对偶,$\tilde{\varphi}=[1,-1]^T,\tilde{\varphi}_1=[0,1]^T$。

(c)以 $\mathbf{R}^2$ 为基础的 $\varphi_0=[1,0]^T,\varphi_1=[-1/2,\sqrt{3}/2]^T$ 和 $\varphi_1=[-1/2,-\sqrt{3}/2]^T$,以及对于 $i=\{0,1,2\}$,它们的对偶 $\tilde{\varphi}_i=2\varphi_i/3$。

提示:必须使用向量内积代替 7.2.1 节中的整数内积。

7.11 说明尺度函数：

$$\varphi(x)=\begin{cases}1 & 0.25 \leqslant x < 0.75 \\ 0 & \text{其他}\end{cases}$$

并未满足多分辨率分析的第二个要求。

7.12 写一个尺度空间 V_3 的表达式作为尺度函数 $\varphi(x)$的一个函数。使用式(7.2.14)定义的哈尔尺度函数，在取 $k=\{0,1,2\}$时画出哈尔 V_3 尺度函数。

★7.13 对于哈尔小波函数画出小波 $\psi_{3,3}(x)$。写出一个关于哈尔尺度函数的$\psi_{3,3}(x)$的表达式。

7.14 假设函数 $f(x)$是哈尔尺度空间 V_3 的一员——即，$f(x)\in V_3$。使用式(7.2.22)将 V_3 表示为一个尺度空间 V_0 和任意需要的小波空间。如果$f(x)$在区间[0,1)的外面为0，大略画出基于你的表达式、函数 $f(x)$的线性展开所要求的尺度函数和小波函数。

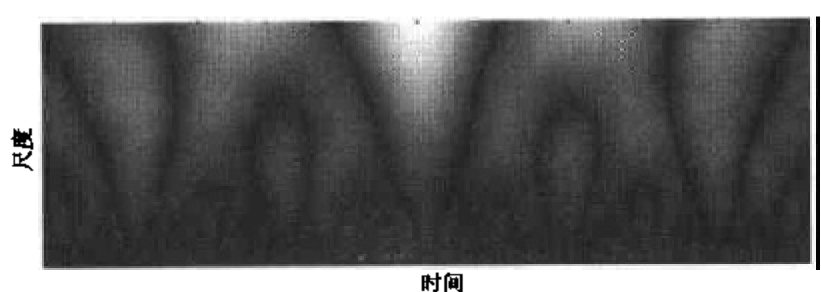
7.15 令起始尺度 $j_0=1$，计算例7.7中使用到的函数的小波级数展开的前四项。写出得到的关于涉及的尺度和小波函数的展开式。在起始尺度为 $j_0=0$ 的地方，得出的结果与例子相比较的结果如何？

7.16 式(7.3.5)和式(7.3.6)中的 DWT 是起始尺度 j_0 的函数。

(a)令 $j_0=1$(而不是0)重新计算例7.8中函数 $f(n)=\{1,4,-3,0\}$在区间 $0\leqslant n\leqslant 3$ 内的一维 DWT。

(b)使用(a)的结果根据变换值计算 $f(1)$。

★7.17 下列以一维函数为基础的连续小波变换揭示了什么？



7.18 (a)习题7.17的连续小波变换是计算机生成的。它所基于的函数首先在离散区间中取样。连续变换是什么？或者说它与函数的离散小波变换有什么区别？

★(b)在什么条件下选择 DWT 比选择 CWT 更好？而 CWT 比 DWT 好的时候，能比 DWT 好几倍吗？

★7.19 画出计算习题7.16中的变换所需要的 FWT 滤波器组。使用适当的序列标记所有的输入和输出。

7.20 一个 M 点快速小波变换的计算复杂度为 $O(M)$。即，操作的次数与 M 成比例。决定比例常数的因素是什么？

7.21 ★(a)如果图7.28(a)的三尺度 FWT 滤波器组的输入是哈尔尺度函数$\varphi(n)=1,n=0,1,\cdots,7$，而 n 取其他值时 $\varphi(n)=0$，就哈尔变换而论，变换结果是什么？

(b)如果输入是对应的哈尔小波函数 $\psi(n)=\{1,1,1,1,-1,-1,-1,-1\},n=0,1,\cdots,7$，那么变换是怎样的？

(c)什么样的输入序列会产生$\{0,0,0,0,0,0,B,0\}$的变换，并具有非零的系数 $W_\psi(2,2)=B$？

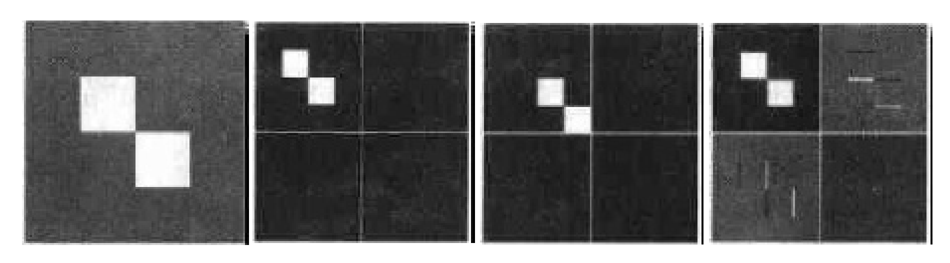
★7.22 二维快速小波变换与 7.2.1 节中的金字塔编码方案相似。它们在哪些地方相似？给定图 7.8(a)中的三尺度小波变换，如何构造一个对应的近似值金字塔？这个金字塔有几级？

7.23 计算关于习题 7.9 中 2×2 图像的哈尔小波的二维小波变换。画出需要的滤波器组并使用适当的阵列标记所有的输入和输出。

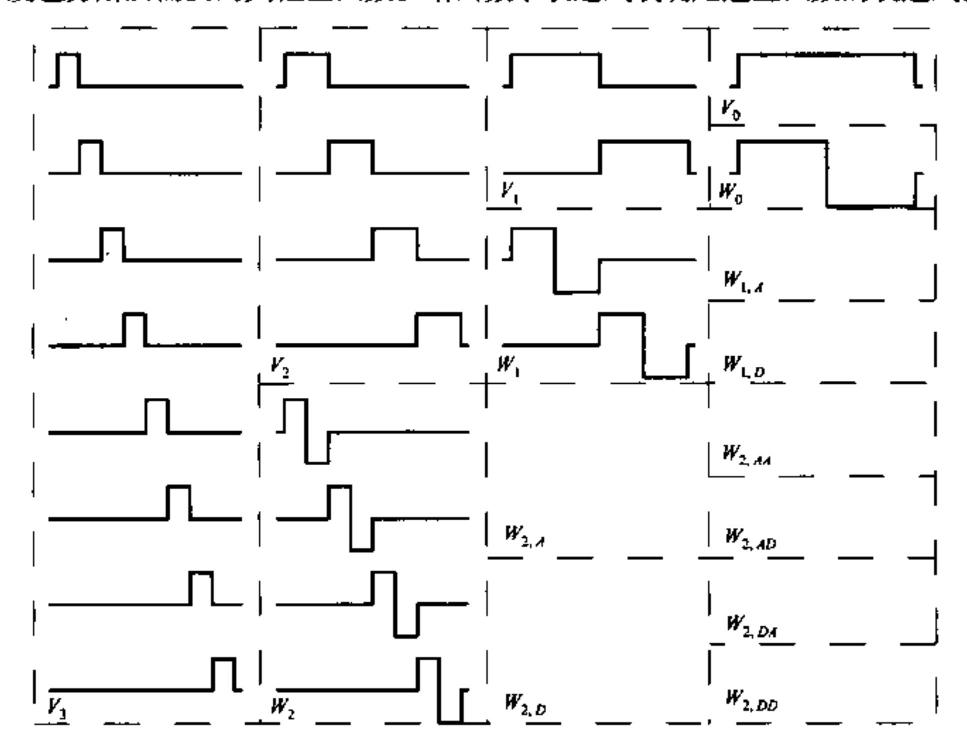
★7.24 在傅里叶域中，

$$f(x-x_0,y-y_0)\Leftrightarrow F(u,v)e^{-2\pi(ux_0/M+vy_0/N)}$$

及平移并不影响$|F(u,v)|$的显示。使用下列的图像序列，解释小波变换的平移特性。最左边的图像在 128×128 大小的灰色背景的中心位置包含两个 32×32 大小的白色方块。第二幅图像(从左边数)是第一幅图像关于哈尔小波的单级小波变换。第三幅图像是在将方块向右下方平移了 32 个像素后，原始图像的小波变换结果。最后一幅图像(最右边的)是原始图像向右下方平移了一个像素之后的小波变换结果。



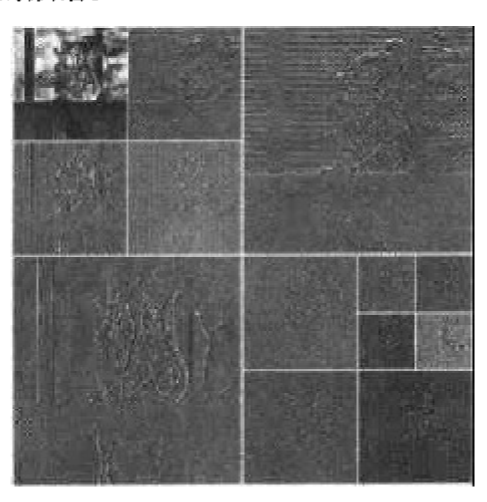
7.25 下列图表说明了一个四尺度快速小波变换的哈尔小波和尺度函数。大略画出完全的三尺度包分解所需要的其他基函数。给出数学表达式或确定这些函数的表达式。



7.26 图 7.1 中的花瓶的一种小波包分解显示如下。

(**a**)画出对应的分解分析树，用适当的尺度和小波空间的名字标记所有节点。

(**b**)画出并标记分解的频谱。



7.27 使用哈尔小波，确定函数 $f(n)=0.25, n=0,1,2,\cdots,15$ 时的最小熵包分解。使用非归一化的香农熵，

$$E[f(n)] = \sum_n f^2(n)\ln[f^2(n)]$$

作为最小化的判断准则。画出最佳树，用计算的熵值标记每个节点。

第8章　图像压缩

每天都有大量的信息用数字进行存储、处理和传送。各个公司通过互联网向合作伙伴、投资商和潜在的消费人群提供金融方面的数据信息、年报、存货目录和产品信息。定单登记和跟踪是两种最基本的网上交易活动。这些每日往来的信息给我们的家庭带来了便利和安逸。美国已经将整个美国国会图书馆的图书(及一些馆藏物品)编制了目录，使其成为世界上最大的电子图书馆，以此作为其进行数字化和建立电子政府的第一步；同时，人们期盼的有线电视点播也即将成为现实。因为网上的许多信息是以图像形式存储的，所以对于存储(见2.4.2节)和通信的需求是无限的。而数据压缩方法比起数据的存储和/或传输具有更为突出的实用价值和商业意义。

图像压缩所解决的问题是尽量减少表示数字图像时需要的数据量。减少数据量的基本原理是除去其中多余的数据。以数学的观点来看，这一过程实际上就是将二维像素阵列变换为一个在统计上无关联的数据集合。这种变换在图像存储或传输之前进行。在以后的某个时候，再对压缩图像进行解压缩来重构原图像或原图像的近似图像。

人们对图像压缩开始感兴趣可以追溯到35年前。最初在这一领域研究的焦点集中在建立一种模拟的方法以便减少视频传输所需的带宽。这一过程称为带宽压缩。数字式计算机的出现和后来先进的集成电路的发展，导致了这方面研究的重点从模拟方式转移到数字压缩方法上来。随着最近几种关键性的国际图像压缩标准的正式采用，这一领域在理论研究的实际应用方面有了重大的发展，这是自20世纪40年代C.E.Shannon和其他人首先提出信息的概率论观点与信息的表达、传输和压缩以来又一重要的发展阶段。

当前，图像压缩被认为是一种“开放技术”。除了刚才提到的一些领域外，由于现代图像传感器不断提高空间分辨率以及电视广播标准的不断发展，图像压缩也成为一种基本技术。再有，数据压缩在许多重要且性质不同的应用领域中扮演着主要角色，比如，电视会议、遥感(使用卫星成像进行天气预报和其他地球资源应用)、记录文献和医疗成像、传真(FAX)、军事上的远程遥控车辆驾驶、空间中的及危险废弃物管理等方面。简而言之，不断扩大的应用领域依赖于对二值图像、灰度图像和彩色图像进行有效的处理、存储和传输。

在本章中，将对图像压缩处理进行理论和实践两方面的研究。8.1节到8.3节是基本原理的介绍，这部分包含了这一学科的全部理论。8.1节讲述了各种图像压缩算法用到的数据冗余概念。在8.2节中介绍用于普通压缩-解压缩处理的以模型为基础的范例。8.3节从信息论的角度在细节上分析一些基本概念以及它们在建立信息表示的基本限制方面的作用。

8.4节到8.6节涵盖了图像压缩实际应用的各个方面，包括实际使用中的主要技术和标准，这些标准有助于该学科领域的扩大与认可。压缩技术又可以分为两个主要的类别：信息保持编码和信息有损编码。8.4节讲解第一类编码的各种方法，它们在编档保存方面很有用处(如用于法律文档和医疗记录保存)。这些方法允许对图像进行无损的压缩和解压缩。8.5节描述了第二类编码的各种方法。这些方法提供更高的数据压缩程度，但重新生成的图像不如原图。有损图像压缩用于广播电视、视频会议和数据传真等方面，在这些应用中，出现一定量

的错误作为提高压缩效能的一种平衡是可接受的。最后，8.6节介绍了现存的和已经提出的各种图像压缩标准。

8.1　基础

术语“数据压缩”指减少表示给定信息量所需的数据量。数据和信息之间必须给予明确的区分。这两个概念的意义是不相同的。实际上，数据是信息传送的手段。对相同数量的信息可以以不同数量的数据表示。比如，有这样一种可能的情况，对于同一个故事可以叙述得冗长啰嗦，也可以说得简明扼要。这里，感兴趣的信息是这个故事；词句是用于表达信息的数据，如果两个不同的人用不同数量的词句讲述同样的故事，那么这个故事就有了两个不同的版本，且至少有一个版本包含了不必要的数据。即，这个版本的故事所包含的数据中(或词句)有与故事无关联的信息或只是重述已经知道了的信息。这就叫做包含了数据冗余。

数据冗余是数字图像压缩的主要问题。它不是一个抽象的概念，而是一个在数学上可以进行量化的实体。如果 n_1 和 n_2 代表两个表示相同信息的数据集合中所携载信息单元的数量，则第一个数据集合(用 n_1 表示的集合)的相对数据冗余 R_D 可以定义为：

$$R_D = 1 - \frac{1}{C_R} \tag{8.1.1}$$

这里 C_R 通常称为压缩率，定义为：

$$C_R = \frac{n_1}{n_2} \tag{8.1.2}$$

对于 $n_2 = n_1$ 的情况，$C_R = 1$，$R_D = 0$，表示(相对于第二个数据集合)信息的第一种表达方式不包含冗余数据。当 $n_2 \ll n_1$ 时，$C_R \to \infty$，$R_D \to 1$，意味着显著的压缩和大量的冗余数据。最后，当 $n_2 \gg n_1$ 时，$C_R \to 0$，$R_D \to -\infty$，表示第二个集合中含有的数据大大超过原表达方式的数据量。当然通常这种数据扩展是不希望出现的情况。通常情况下，C_R 和 R_D 分别在开区间 $(0, \infty)$ 和 $(-\infty, 1)$ 内取值。比较实际的压缩率，比如 10(或者说 10:1)，意味着对应第二个集合或压缩过的数据集合的每一个单元，第一个数据集合中有 10 个信息携载单元(比如，比特)。相应的冗余度为 0.9，这表示在第一个数据集合中有 90% 的数据是冗余数据。

在数字图像压缩中，可以确定三种基本的数据冗余并加以利用：编码冗余、像素间冗余和心理视觉冗余。当这三种冗余中的一种或多种得到减少或消除时，就实现了数据压缩。

8.1.1　编码冗余

在第 3 章中，探讨了在假设图像的灰度级为随机量的基础上通过直方图处理进行图像增强的技术。说明了可以通过图像灰度级的直方图得到关于图像外观的大量信息。在本节中，将利用相似的表示方法说明如何通过图像的灰度级直方图深入了解编码结构①，从而减少表达图像所需的数据量。

① 编码是符号系统(字符、数字、位及类似符号)，它用于表示信息的主体或事件的集合。每个信息或事件都被赋予一个编码符号序列，称为码字。每个码字中符号的个数是这个码字的长度。最著名的代码之一是 Paul Revere 1775 年 4 月 18 日使用的。

这里再一次假设区间[0,1]内的一个离散随机变量 r_k 表示图像的灰度级,并且每个 r_k 出现的概率为$p_r(r_k)$:和第3章中一样:

$$p_r(r_k) = \frac{n_k}{n} \quad k = 0,1,2,\cdots,L-1 \tag{8.1.3}$$

这里 L 是灰度级数,n_k 是第k 个灰度级在图像中出现的次数,n 是图像中的像素总数。如果用于表示每个 r_k 值的比特数为$l(r_k)$,则表达每个像素所需的平均比特数为:

$$L_{avg} = \sum_{k=0}^{L-1} l(r_k)p_r(r_k) \tag{8.1.4}$$

就是说,将表示每个灰度级值所用的比特数和灰度级出现的概率相乘,将所得乘积相加后得到不同灰度级值的平均码字长度。因此,对 $M \times N$ 大小的图像进行编码所需的比特数为 MNL_{avg}。

使用 m 比特自然(此处,由图像中灰度级数目决定)二进制编码[①] 表示图像的灰渡级将式(8.1.4)右边减少到 m 比特。就是说,当使用 m 来替代$l(r_k)$时,$L_{avg} = m$。然后,常数 m 可以提到和式的外面,只剩下 $p_r(r_k)$在$0 \leqslant k \leqslant L-1$ 时的和,当然,其结果为1。

例 8.1 变长编码的说明

表8.1显示了一幅具有8个灰度级的图像的灰度级分布。如果其3比特自然二进制编码[见表8.1中的编码1和$l_1(r_k)$]用于表示8个可能的灰度级,则 L_{avg}为3比特,因为对所有 r_k,$l_1(r_k)=3$ 比特。如果使用表8.1中的编码2,表示图像所需编码的平均比特数就减少为:

$$\begin{aligned} L_{avg} &= \sum_{k=0}^{7} l_2(r_k)p_r(r_k) \\ &= 2(0.19)+2(0.25)+2(0.21)+3(0.16)+4(0.08)+ \\ &\quad 5(0.06)+6(0.03)+6(0.02) \\ &= 2.7 \text{ 比特} \end{aligned}$$

表 8.1 变长编码的例子

r_k	$p_r(r_k)$	编码1	$l_1(r_k)$	编码2	$l_2(r_k)$
$r_0=0$	0.19	000	3	11	2
$r_1=1/7$	0.25	001	3	01	2
$r_2=2/7$	0.21	010	3	10	2
$r_3=3/7$	0.16	011	3	001	3
$r_4=4/7$	0.08	100	3	0001	4
$r_5=5/7$	0.06	101	3	00001	5
$r_6=6/7$	0.03	110	3	000001	6
$r_7=1$	0.02	111	3	000000	6

① 自然二进制编码是,每个事件或每段信息都可以用来自 m 比特二进制计数序列的 2^m 个 m 比特二进制码的其中一个来表示。

根据式(8.1.2),得到的压缩率是 3/2.7 或 1.11。因此使用编码 1 有大约 10% 的数据是冗余的。根据式(8.1.1)可以确定准确的冗余水平:

$$R_D = 1 - \frac{1}{1.11} = 0.099$$

图 8.1 用编码 2 说明了实现压缩的基本原理。它显示了图像的直方图[$p_r(r_k)$与 r_k 的曲线]和 $l_2(r_k)$。因为这两个函数成反比——就是说,当 $p_r(r_k)$减少时,$l_2(r_k)$增加——编码 2 中最短的码字赋予图像中出现频率最高的灰度级。

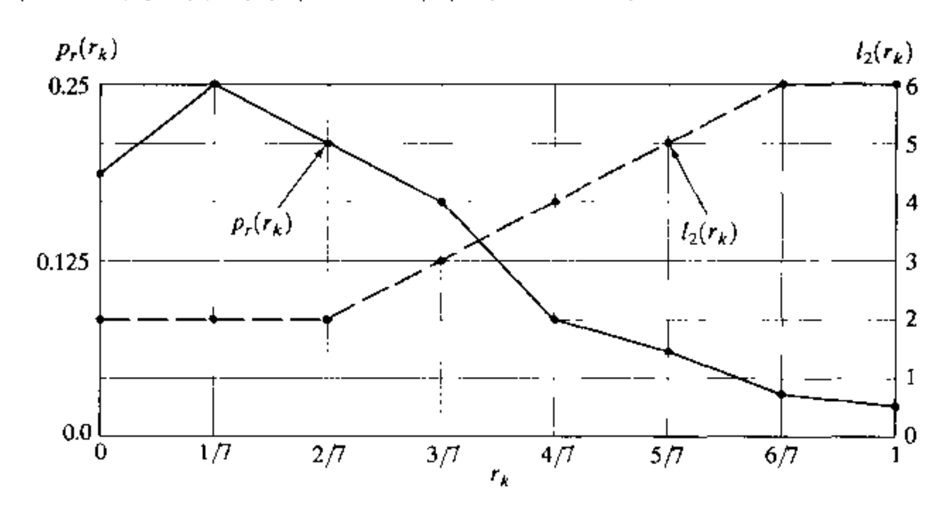


图 8.1 用变长编码的数据压缩基本原理的图表表示

在前述的例子中,通过用尽量少的比特数表达尽可能多的灰度级以实现数据的压缩。这种处理通常称为变长编码。如果图像的灰度级在编码时用的编码符号数多于表示每个灰度级实际所需的符号数[就是说,这个编码无法使式(8.1.4)得到最小值],则用这种编码得到的图像包含编码冗余。通常,当被赋予事件集的编码(比如灰度级值)如果没有充分利用各种结果出现的概率去选择,就会存在编码冗余。当一幅图像的灰度级直接用自然二进制编码来表示时,冗余总会存在。在这种情况下,处理编码冗余的根本基础就是:图像是由具有规则的、在某种程度上具有可预测的形态(形状)和反射的对象组成的,并且通常对图像进行取样以便描述的对象远大于图像元素。在大多数图像中,正常情况下的结果是某个灰度级比其他灰度级有更大的出现可能性(就是说,多数图像的直方图是不均匀的)。它们灰度级的自然二进制编码对有最大和最小可能性的值分配相同的比特数,因此无法得到式(8.1.4)的最小值,产生了编码冗余。

8.1.2 像素间冗余

考虑图 8.2(a)和(b)中所示的图像。如图 8.2(c)和(d)所示,这些图像实质上有同样的直方图。同时注意这两幅直方图都是具有三个波峰的,表明灰度级存在三个主要的值域。因为这些图像中的灰度级的出现概率不是等可能性的,所以可以使用变长编码减少由于对像素进行统一长度的编码或自然二进制编码带来的编码冗余。然而,编码处理不会改变图像的像素之间相关性的级别。换句话说,用于表示每幅图像的灰度级的编码与像素之间的相关性无关。这些相关来自于图像中对象之间的结构或几何关系。

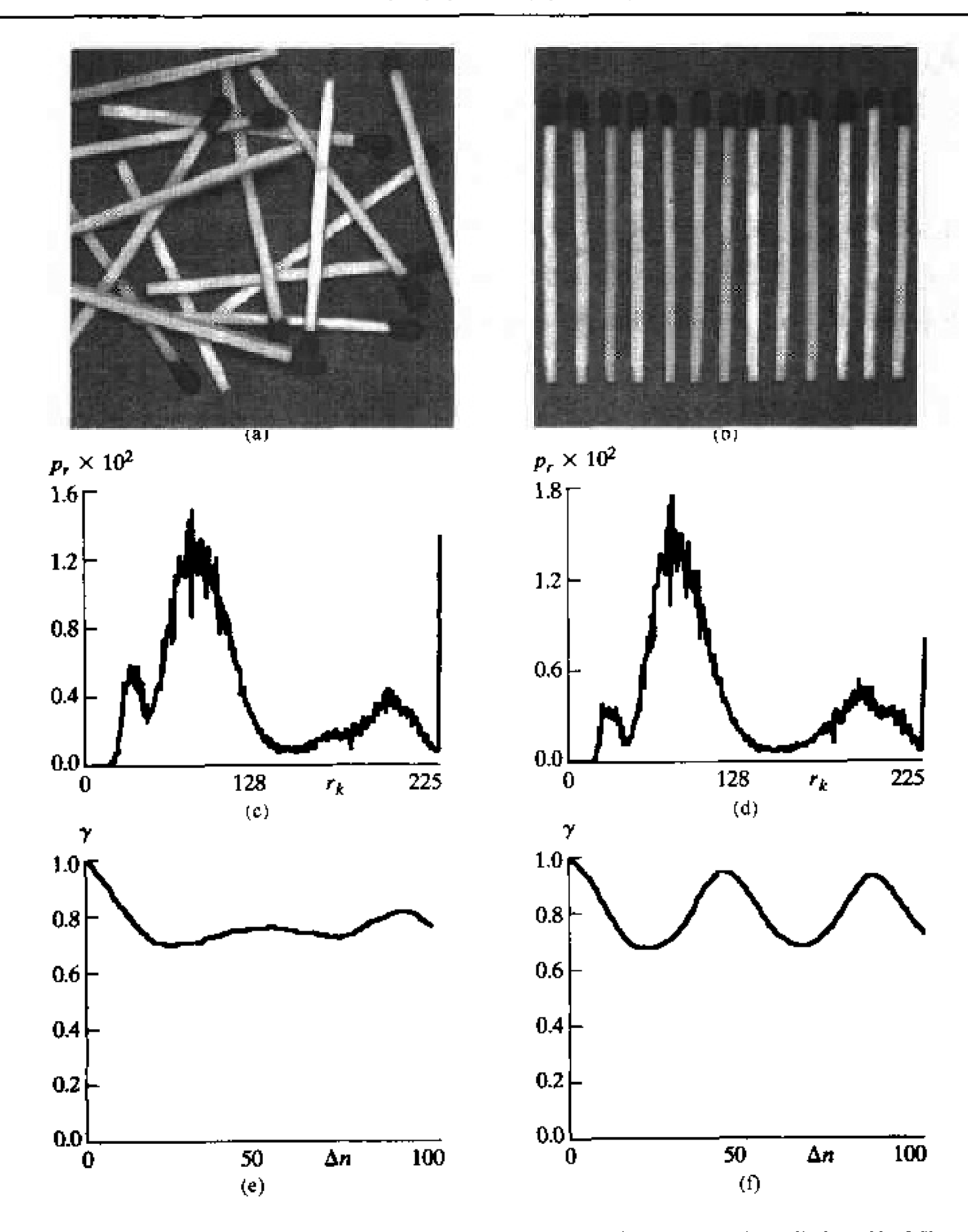


图 8.2　两幅图像和它们的灰度级直方图以及沿着某条线计算的归一化自相关系数

图 8.2(e)和(f)显示了沿着每幅图像的某条线计算得到的各自的自相关系数。这些系数是使用式(4.6.30)的归一化版本计算得到的:

$$\gamma(\Delta n)=\frac{A(\Delta n)}{A(0)} \tag{8.1.5}$$

这里,

$$A(\Delta n)=\frac{1}{N-\Delta n}\sum_{y=0}^{N-1-\Delta n} f(x,y)f(x,y+\Delta n) \tag{8.1.6}$$

式(8.1.6)中的比例因子说明了 Δn 取整数值时,求和项变化的项数。当然,Δn 必须严格小于线上的像素数目 N。变量 x 是计算中所用的线的坐标。注意图 8.2(e)和(f)中显示的函数形

状之间戏剧性的差异。它们形状上的差异可以定性地同图 8.2(a)和(b)联系起来。这种关系在图 8.2(f)中是显而易见的,这里被 45 和 90 个样本分开的像素之间的高度相关可以直接同图 8.2(b)中垂直方向的火柴之间的间隔联系起来。另外,两幅图像中相邻的像素具有高度相关性。当 Δn 为 1 时,对于图 8.2(a)和(b)的图像,γ 分别是 0.9922 和 0.9928。这些值大多数是对电视图像进行适当的取样得到的典型值。

这些说明反映了另一种数据冗余的重要形式——一种图像中与像素间相关的直接关系。因为任何给定像素的值可以根据与这个像素相邻的像素进行适当的预测，所以由单个像素携载的信息相对较少。单一像素对于一幅图像的多数视觉贡献是多余的；它的值可以通过以与其相邻的像素值为基础进行推测。许多命名，包括空间冗余、几何冗余和帧间冗余都用来表示这些像素间的依赖性。使用术语“像素间冗余”来代表上述所有的名称。

为了减少图像中的像素间冗余，通常用于人类观察和解释的二维像素阵列必须变换为更有效的形式（但通常是“不可见的”）。例如，相邻像素之间的差异可以用来描绘图像。这种类型的变换（确切地说是消除像素间冗余）被认为是映射。如果初始图像元素可以根据经过变换的数据集合进行重构，则称这些映射为可逆映射。

例 8.2　行程编码的一个简单说明

图 8.3 列举了一个简单的映射过程。图 8.3(a)描绘了一个 1 in × 3 in 大小,以大约 330 dpi (每英寸点数)的取样率绘制的电子部件的某部分。图 8.3(b)显示了这幅图像的二值图,图 8.3(c)描绘了图像中的一条线的灰度剖面和用于得到二值图像的门限(见 3.1 节)。因为二值图像包含很多亮度为常数值的区域,所以沿着每条扫描线 $f(x,0),f(x,1),\cdots,f(x,N-1)$从像素到数对序列$(g_1,w_1),(g_2,w_2),\cdots$的映射是更为有效的表达方法。这里 g_i 表示沿着线遇到的第 i 个灰度级,w_i 表示第 i 次扫描的行程。换句话说,经门限处理的图像使用图像的恒定灰度级行程(一种不可见表示)的长度和值比用二维二值像素阵列表达更为有效。

图 8.3(d)显示了行程对应于图 8.3(c)的门限线剖面进行编码的结果。表示 1 024 比特的二值数据只需 88 比特。实际上,图 8.3(b)中整个 1 024 × 343 大小的部件可以减少为 12 166行程。由于表示每个行程对需要 11 比特,所以得到的压缩率和相应的相对冗余为:

$$C_R = \frac{(1\,024)(343)(1)}{(12\,166)(11)} = 2.63$$

且

$$R_D = 1 - \frac{1}{2.63} = 0.62$$

8.1.3　心理视觉冗余

在 2.1 节中注意到人眼感觉到的区域的亮度不仅仅取决于该区域的反射光,还取决于其他一些因素。例如,在亮度不变的区域中同样可以感觉到亮度的变化(马赫带)。产生这种现象的原因是由于眼睛对所有视觉信息感受的灵敏度不同。在正常的视觉处理过程中各种信息的相对重要程度不同。那些不十分重要的信息称做心理视觉冗余。这些冗余在不会削弱图像感知质量的情况下可以消除。

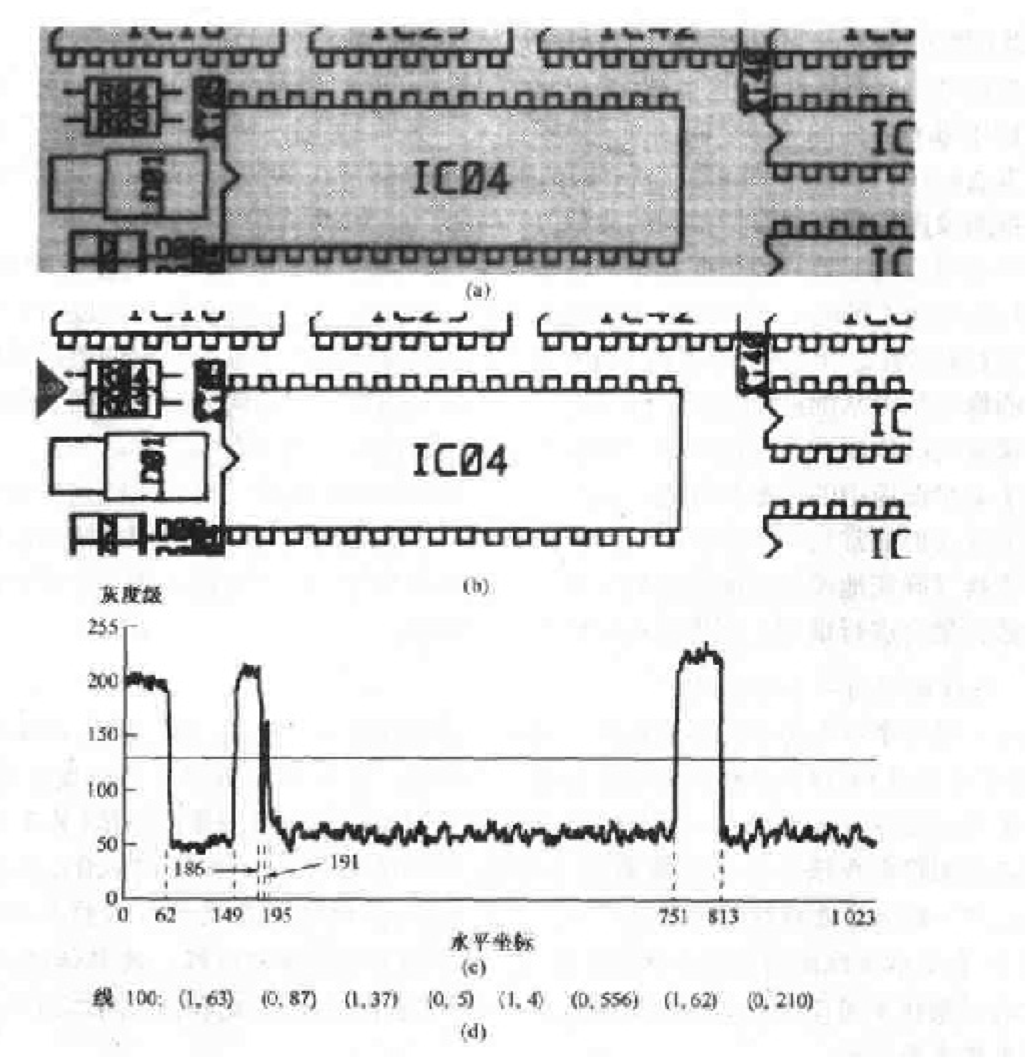


图 8.3　行程编码的说明。(a)原图,(b)标记了线 100 的二值图像,(c)线状剖面和二值化门限,(d)行程编码

不要对存在心理视觉冗余而感到奇怪,因为人类对图像信息的感知并不涉及对图像中每个像素值的定量分析。通常,观察者寻找可区别的特征,比如边线或纹理区域,然后在心里将它们合并成可识别的组群。然后,人类的大脑将这些组群与先前已有的知识相联系以便完成图像的解释过程。

心理视觉冗余来自先前讨论到的冗余的基本差异。与编码冗余和像素间冗余不同,心理视觉冗余是与真实的或可定量的视觉信息相联系的。因为对于正常的视觉处理过程,信息本身不是本质,所以消除心理视觉冗余是可能的。由于消除心理视觉冗余数据会导致一定量信息的丢失,所以这一过程通常称为"量化"。这个术语与词句的通常用法是相一致的,它表示从一个范围很广的输入值集合到有限的输出值的映射,如 2.4 节中讨论的那样。由于这种映射是不可逆的(视觉信息有损失),量化的结果导致了数据有损压缩。

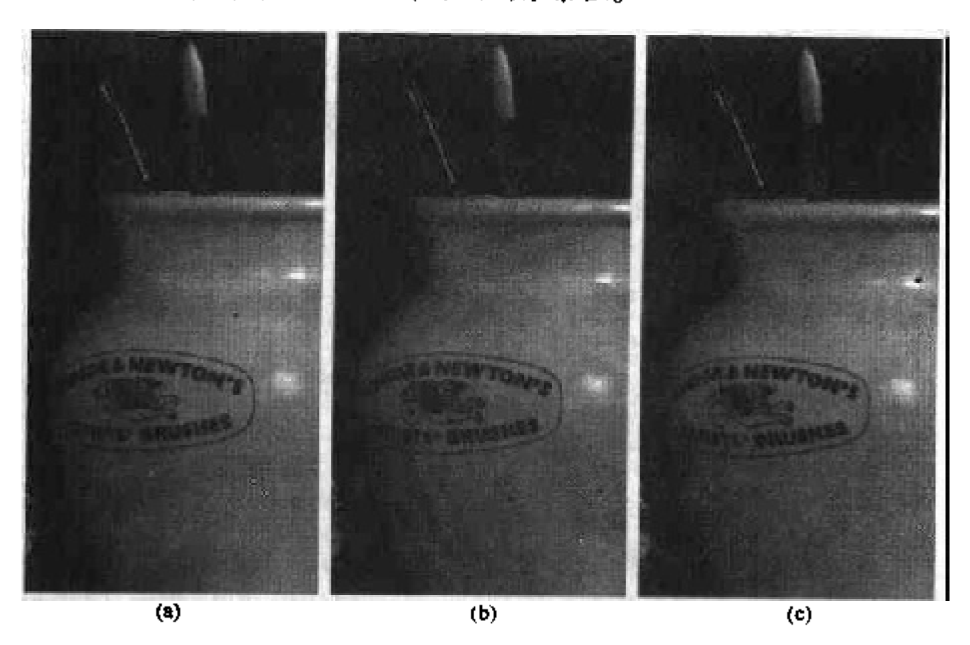
例 8.3　通过量化进行压缩

考虑图 8.4 中的图像。图 8.4(a)显示了一幅具有 256 个灰度级的单色图像。图 8.4(b)显示了将这幅图像均匀量化为 4 比特或 16 种可能的灰度级后的结果。压缩率为 2:1。注

意，和 2.4 节中讨论的一样，在原图像先前平滑的区域出现了假轮廓。这是对图像灰度级进行粗略表示所带来的固有视觉影响。

图 8.4(c)说明充分利用人类视觉系统特性的量化过程可以大大改善对图像的表现效果。尽管这种量化过程的压缩率还是只有 2:1，为减少假轮廓而增加了额外的开销，但减少了讨厌的颗粒状纹路。产生这种结果所使用的方法就是改进的灰度级(IGS)量化方法。它利用眼睛对边缘固有的敏感性，通过将一个伪随机数加到每个像素上将这些边缘拆散。这个伪随机数是在对结果进行量化之前，根据表示相邻像素灰度级的低位生成的。由于低位完全是随机的(见 3.2.4 节中的位平面)，所以，这样做等于增加了通常与伪轮廓相关的人工边缘随机性的灰度级。而这种随机性取决于图像的局部特征。

表 8.2 对这种方法进行了说明。先由当前的 8 位灰度级值和前面产生的 4 位最低有效位构成一个初始值为零的和。如果当前值的 4 位最高有效位为 1111_2，则用 0000_2 与其相加。将得到的和的 4 个最高有效位的值作为编码像素值。



(a) (b) (c)

图 8.4 (a)原图，(b)均匀量化为 16 个灰度级，(c)用 IGS 量化为 16 个灰度级

表 8.2 IGS 量化过程

像素	灰度级	和	IGS 编码
$i-1$	N/A	00000000	N/A
i	01101100	01101100	0110
$i+1$	10001011	10010111	1001
$i+2$	10000111	10001110	1000
$i+3$	11110100	11110100	1111

改进的灰度级量化是典型的大批量量化过程。这个过程直接对被压缩图像的灰度级进行。量化的过程会导致图像的空间和/或灰度级解析度的降低。得到的伪轮廓或其他相关的影响必须通过使用试探技术补偿量化的视觉影响。例如,用于商业电视广播的标准 2:1 隔行方法就是一种量化形式,这种形式中相邻帧的交错部分在图像视觉品质下降很少的情况下,允许减少视频扫描率。

8.1.4 保真度准则

正如前面所提到的,消除心理视觉冗余数据导致真实的或一定数量的视觉信息的丢失。因为可能会由此失去重要的信息,所以迫切需要一种可重复或可再生的对于丢失信息的性质和范围定量的方法。作为这种评估基础的两类准则是:(1)客观保真度准则和(2)主观保真度准则。

当信息损失的程度可以表示成初始图像或输入图像以及先被压缩而后被解压缩的输出图像的函数时,就说这个函数是基于客观保真度准则的。输入图像和输出图像之间的均方根(rms)误差就是最好的例子。令 $f(x,y)$表示输入图像,并令 $\hat{f}(x,y)$表示由对输入先压缩后解压缩得到的 $f(x,y)$的估计量或近似量。对 x 和 y 的所有值,$f(x,y)$和 $\hat{f}(x,y)$之间的误差 $e(x,y)$可以定义为:

$$e(x,y) = \hat{f}(x,y) - f(x,y) \tag{8.1.7}$$

从而使两幅图像的总体误差为:

$$\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[\hat{f}(x,y) - f(x,y)]$$

这里图像大小为 $M \times N$ 像素。$f(x,y)$和 $\hat{f}(x,y)$之间的均方根误差 e_{rms}则成为在 $M \times N$ 像素上平方误差平均值的平方根,或写为:

$$e_{rms} = \left[\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[\hat{f}(x,y) - f(x,y)]^2\right]^{1/2} \tag{8.1.8}$$

一种关系更为紧密的客观保真度准则是压缩-解压缩图像的均方信噪比。如果认为 $\hat{f}(x,y)$是初始图像和噪声信号 $e(x,y)$的和,则输出图像的均方信噪比用 SNR_{ms}表示,定义为:

$$SNR_{ms} = \frac{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\hat{f}(x,y)^2}{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[\hat{f}(x,y) - f(x,y)]^2} \tag{8.1.9}$$

信噪比的 rms 值 SNR_{rms}通过求取式(8.1.9)的平方根得到。

尽管客观保真度准则提供了一种简单便捷的评估信息损失的机理,但大部分解压缩图像最终还是由人来进行观察的。所以,使用观察者的主观评估衡量图像品质通常是更为适当的。主观评估是通过向典型的观察者显示典型的解压缩图像并将他们的评估结果进行平均得到的。评估可能采取绝对等级或并排对比 $f(x,y)$和 $\hat{f}(x,y)$的形式。表 8.3 显示了一个可行的绝对等级量表。并排对比可以使用诸如{-3,-2,-1,0,1,2,3}这样的标度分别表示主观评估{非常恶劣、恶劣、稍坏、普通、稍好、更好、非常好}。不管使用何种形式,这些评估都称为基于主观保真度准则。

例 8.4 图像品质的对比

图 8.4(b)和(c)的量化图像的 rms 误差分别为 6.93 和 6.78。对应的 rms 信噪比为 10.25 和 10.39。尽管这些值很相近，但两幅编码图像视觉品质的主观评估也许会是，对图 8.4(b)中图像为“勉强可以”等级，对图 8.4(c)中图像为“过得去”等级。

表 8.3 电视图像的等级量表(Frendendall 和 Behrend)

值	等级	描述
1	极好	具有极高品质的图像，和希望的一样好
2	好	高品质的图像，感觉良好，其中的干扰可以接受
3	过得去	具有可接受的品质。其中的干扰不是不可接受的
4	勉强可以	品质不良的图像；希望能得到改进。干扰在某种程度上难于接受
5	差	非常不好的图像，但还可以看。有明显不能接受的干扰
6	不可用	差到无法观看的图像

8.2 图像压缩模型

在 8.1 节中，单独讨论了用于减少或压缩表示一幅图像所要求的数据的三种一般技术。然而，这些技术一般相互结合用于构成实际的图像压缩系统。在本小节中，对这种系统的特点进行全面的分析并建立一种表示它的一般模型。

如图 8.5 所示，一个图像压缩系统包括两个不同的结构块：一个编码器和一个解码器[①]。图像 $f(x,y)$ 输入到编码器中。这个编码器可以根据输入数据生成一组符号。在通过信道进行传输之后，将经过编码的表达符号送入解码器，经过重构后，就生成了输出图像 $\hat{f}(x,y)$。一般来讲，$\hat{f}(x,y)$ 可能是也可能不是原图像 $f(x,y)$ 的准确复制品。如果输出图像是输入的准确复制，系统就是无误差的或具有信息保持编码的系统；如果不是，则在重建图像中就会呈现某种程度的失真。

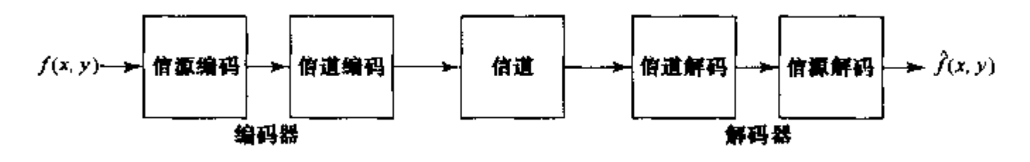


图 8.5 一个常用的图像压缩系统模型

图 8.5 中显示的编码器和解码器都包含两个彼此相关的函数或子块。编码器由一个消除输入冗余的信源编码器和一个用于增强信源编码器输出的噪声抗扰性的信道编码器构成。如预想的那样，一个解码器包括一个信道解码器，它后面跟着一个信源解码器。如果编码器和解码器之间的信道是无噪声的(趋向于无误差)，则信道编码器和信道解码器可以略去，而一般的编码器和解码器分别是信源编码器和信源解码器。

① 称这些块为“压缩器”和“解压缩器”也是合理的。编码器和解码器的称谓反映了信息论对图像压缩领域的影响(将在 8.3 节中进行讨论)。

8.2.1 信源编码器和信源解码器

信源编码器的任务是减少或消除输入图像中的编码冗余、像素间冗余或心理视觉冗余。特定的应用和与之相联系的保真度要求规定了在给定情况下使用的最佳编码方法。通常，这种方法可以通过一系列三种独立的操作建立模型。如图 8.6(a)所示，每种操作被设计用于减少 8.1 节中谈到的三种冗余中的一种。图 8.6(b)描述了对应的信源解码器。

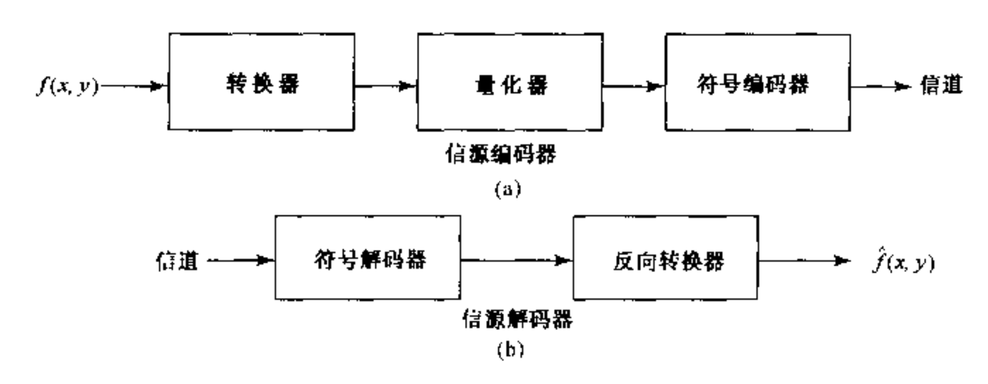


图 8.6 (a)信源编码器和(b)信源解码器模型

在信源编码处理的第一阶段，转换器将输入数据转换为可以减少输入图像中像素间冗余的格式(通常不可见)。这步操作通常是可逆的并且有可能直接减少表示图像的数据量。行程编码(8.1.2 节和 8.4.3 节)就是在整个信源处理的初始阶段对数据进行压缩转换的例子。用一组变换系数(8.5.2 节)表示图像的方法是与此相反的例子。这里，转换程序将图像变换为一个系数阵列。这样做使编码处理的后续阶段中更容易找到像素间冗余以便进行压缩。

在第二阶段中，或图 8.6(a)中的量化器模块中，将转换程序的输出精度调整到与预设的保真度准则相一致。这一步减少了输入图像的心理视觉冗余。如在 8.1.3 节中提到的，这一步的操作是不可逆的。因此，当希望进行无误差压缩时，这一步必须略去。

在第三阶段，即信源编码处理的最后阶段，符号编码器生成一个固定的或可变长编码用于表示量化器输出并将输出转换为与编码相一致。术语“符号编码器”与整个信源编码处理过程是不同的。在大多数情况下，变长编码用于表示经过转换和量化的数据集合。它用最短的码字表示出现频率最高的输出值，以此减少编码冗余。当然，这种操作是可逆的。在符号编码的完成阶段，输入图像经过处理已经消除了 8.1 节中所说的三种冗余。

图 8.6(a)显示了信源编码处理三个相继的操作，但并不是每个图像压缩系统都必须包含这三种操作。比如，回想一下，当希望进行无误差压缩时，必须去掉量化器。另外，某些图像压缩技术是将图 8.6(a)中物理分开的某些模块进行合并后建立的模型。例如，在 8.5.1 节提到的图像压缩系统中，转换器和量化器经常用一个模块表示。这个模块同时执行这两部分的操作。

图 8.6(b)中显示的信源解码器仅包含两部分：一个符号解码器和一个反向转换器。这些模块的运行次序与编码器的符号编码器和转换模块的操作次序相反。因为量化过程导致了不可逆的信息损失，反向量化器模块不包含在图 8.6(b)所示的通常的信源解码器模型中。

8.2.2 信道编码器和解码器

当图8.5中显示的信道带有噪声或易于出现错误时，信道编码器和解码器就在整个译码解码处理中扮演了重要的角色。信道编码器和解码器通过向信源编码数据中插入预制的冗余数据来减少信道噪声的影响。由于信源编码器几乎不包含冗余，所以如果没有附加这种"预制的冗余"，它对噪声传送会有很高的敏感性。

最有用的一种信道编码技术是由R.W.Hamming(Hamming[1950])提出的。这种技术是基于这样的思想，即向被编码数据中加入足够的位数以确保可用的码字间变化的位数最小。例如，Hamming说明如果将3位冗余码加到4位字上，使得任意两个有效码字间的距离[①] 为3，则所有的一位错误都可以检测出来并得到纠正(通过附加额外的冗余位，多位错误也可以得到检测和纠正)。与4位二进制数 $b_3b_2b_1b_0$ 相联系的7位Hamming(7,4)码字 $h_1h_2\cdots h_5h_6h_7$ 是：

$$
\begin{aligned}
h_1 &= b_3 \oplus b_2 \oplus b_0 & h_3 &= b_3 \\
h_2 &= b_3 \oplus b_1 \oplus b_0 & h_5 &= b_2 \\
h_4 &= b_2 \oplus b_1 \oplus b_0 & h_6 &= b_1 \\
& & h_7 &= b_0
\end{aligned}
\tag{8.2.1}
$$

这里+表示异或运算。注意 h_1，h_2 和 h_4 位分别是位字段 $b_3b_2b_0$，$b_3b_1b_0$ 和 $b_2b_1b_0$ 的偶校验位(回想一下，如果值为1的位数是偶数，则二进制位串具有偶校验)。

为了将汉明(Hamming)编码结果进行解码，信道解码器必须为先前设立的偶校验的各个位字段进行奇校验并检查译码值。一位错误由一个非零奇偶校验字 $c_4c_2c_1$ 给出，这里，

$$
\begin{aligned}
c_1 &= h_1 \oplus h_3 \oplus h_5 \oplus h_7 \\
c_2 &= h_2 \oplus h_3 \oplus h_6 \oplus h_7 \\
c_4 &= h_4 \oplus h_5 \oplus h_6 \oplus h_7
\end{aligned}
\tag{8.2.2}
$$

如果找到一个非零值，则解码器只需简单地在校验字指出的位置补充码字比特。解码的二进制值 $h_3h_5h_6h_7$ 就从纠正后的码字中提取出来。

例8.5 汉明编码

考虑在带有噪声的信道上传送表8.2中的4位IGS数据。一个一位的错误会导致解压缩像素与正确的值有最多128个灰度级[②] 的偏差。一个汉明信道编码器可以通过加入足够的冗余来检测和纠正一位错误来增加信源编码IGS数据的抗扰度。根据式(8.2.1)，表8.2中第一个IGS值的汉明编码值是 1100110_2。因为汉明信道编码器将表示IGS值所需的位数从4位增加到7位，在IGS的例子中提到的2:1的压缩率减少到8/7或1.14:1。压缩率的降低是增强噪声抗扰度所付出的代价。

---

① 两个码字之间的距离定义为：为了得到另一个码字，其中一个码字中的数字必须变化的最少数目。例如，101101和011101之间的距离是2。码字间的最小距离是任意两个码字之间相异数字的最少数目。

② 一个用于解码4位IGS数据的简单程序是用16乘以IGS值的十进制量。例如，如果IGS值是1110，解码后的灰度级为14×16或224。如果这个IGS值的最高位被错误地传送为0，解码后的灰度级就变成96。得到的误差为128个灰度级。

8.3 信息论要素

在 8.1 节中介绍了几种减少表示图像的数据量的方法。这使我们很自然地联想到一个问题:显示一幅图像究竟需要多大的数据量? 就是说,有没有满足充分描述一幅图像且没有信息丢失的最小数据量呢? 信息论提供了回答这个问题以及与此相关问题的理论框架。

8.3.1 测量信息

信息论的基本前提是信息的产生可以被模拟为一个概率过程,这个过程可以用与我们的直觉相一致的方法度量。依照这个假设,发生概率为 $P(E)$的随机事件 E 被认为是包含:

$$I(E) = \log \frac{1}{P(E)} = -\log P(E) \tag{8.3.1}$$

的信息单元。量 $I(E)$通常被称为 E 的自信息。一般来讲,归因于事件 E 的自信息量与E 的概率的关系是反向的。如果 $P(E)=1$(即,事件总会发生),则 $I(E)=0$,没有归因于它的信息。就是说,因为事件没有不确定性,所以在有事件发生的通信过程中,不会传递任何信息。然而,如果 $P(E)=0.99$,事件 E 已经发生的通信传输少量信息。事件 E 没有发生的通信会传递更多的信息,因为这个输出的可能性很小。

式(8.3.1)中对数的底数决定了衡量信息①的单位。如果使用 m 作为对数的底数,则这种度量被称为是 m 元单位。如果底数选择为 2,信息的结果单元是比特。注意,如果 $P(E)=1/2$,$I(E)=-\log_2 1/2$或 1 比特。就是说,当两种出现概率相等的事件发生时,传送的信息量是 1 比特。这类情况的简单例子就是抛硬币并传送结果。

8.3.2 信息信道

当自信息在信息源和信息用户之间传送时,信息源称为与信息用户之间通过信息信道相连。信息信道是连接信源和用户的物理媒介。它可以是电话线、电磁能量传播路径或数字计算机中的导线。图 8.7 显示了一个离散信息系统的简单数学模型。这里,系统的容量是特别重要的参量,定义为系统传输信息的能力。

假设图 8.7 中的信息源根据一个有限的或无限但可数的符号集合生成一个任意的符号序列。就是说,信源的输出是离散的随机变量。信源符号集$\{a_1,a_2,\cdots,a_J\}$称为信源字母表 A,集合中的元素表示为 a_j,称为符号或字符。信源产生符号 a_j 的事件概率是$P(a_j)$,且有:

$$\sum_{j=1}^{J} P(a_j) = 1 \tag{8.3.2}$$

一个 $J\times 1$ 向量 $\mathbf{z}=[P(a_1),P(a_2),\cdots,P(a_J)]^T$ 通常用于表示所有信源符号的概率集合$\{P(a_1),P(a_2),\cdots,P(a_J)\}$。有限总体集合$(A,\mathbf{z})$完全描述了信息源。

离散信源产生符号 a_j 的概率是 $P(a_j)$,因此生成一个单一信源符号产生的自信息根据式(8.3.1)应是 $I(a_j)=-\log P(a_j)$。如果产生了 k 个信源符号,则大数定律保证对于一个充

① 当没有明确指定表达式中 log 的底数时,可以将它解释为任意底数及对应的信息单元。

分大的 k 值，符号 a_j 将(平均)被输出 $kP(a_j)$次。因此，根据 k 输出得到的平均自信息是：

$$-kP(a_1)\log P(a_1)-kP(a_2)\log P(a_2)-\cdots-kP(a_J)kP(a_J)$$

或

$$-k\sum_{j=1}^{J}P(a_j)\log P(a_j)$$

每个信源输出的平均信息，表示为 $H(\mathbf{z})$：

$$H(\mathbf{z})=-\sum_{j=1}^{J}P(a_j)\log P(a_j) \tag{8.3.3}$$

这个量被称为信源的不定度或熵。它定义了观察单个信源输出得到的平均信息量(每个符号的 m 元单元)。随着这个量的增加，信源具有更大的不定度且因此有更多的信息与信源相联系。如果信源符号的出现是等可能性的，则式(8.3.3)的熵或不定度被最大化而且信源对每个信源符号提供可能的最大信息量。

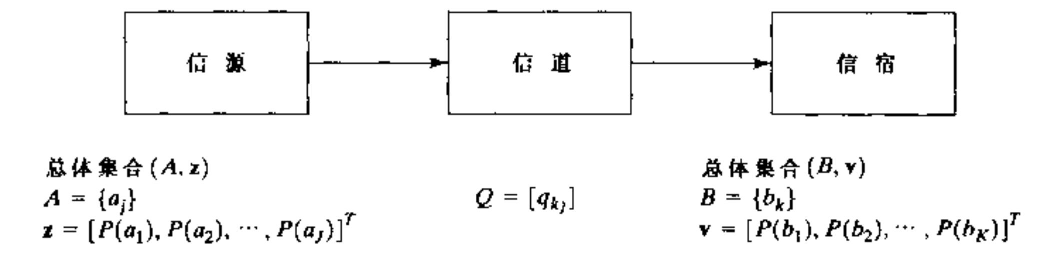


图 8.7　一个简单的信息系统

经过对信源的模型化，更容易建立信道的输入输出特征。因为将图 8.7 中信道的输入模型化为离散随机变量，所以传送到信道输出的信息也是离散随机变量。如信源随机变量一样，这个随机变量的值也是来自一个被称为信道字母表 B 的有限或可数的无限符号集合 $\{b_1,b_2,\cdots,b_K\}$。提交给信息用户的字符 b_k 的概率是 $P(b_k)$。有限的集合$(B,\mathbf{v})$完整地描述了信道输出和用户收到的信息，这里 $\mathbf{v}=[P(b_1),P(b_2),\cdots,P(b_K)]^T$。

给定信道输出的概率 $P(b_k)$和信源 $\mathbf{z}$ 的概率分布由下式相联系：①

$$P(b_k)=\sum_{j=1}^{J}P(b_k\mid a_j)P(a_j) \tag{8.3.4}$$

这里 $P(b_k\mid a_j)$是给出信源符号 a_j，被接收的输出符号 b_k 的条件概率。如果根据式(8.3.4)，得到的条件概率被排列成 $K\times J$ 的矩阵 $\mathbf{Q}$：

$$\mathbf{Q}=\begin{bmatrix}P(b_1\mid a_1) & P(b_1\mid a_2) & \cdots & P(b_1\mid a_J)\\ P(b_2\mid a_1) & \cdot & \cdots & \cdot\\ \vdots & \vdots & \cdots & \vdots\\ P(b_K\mid a_1) & P(b_K\mid a_2) & \cdots & P(b_K\mid a_J)\end{bmatrix} \tag{8.3.5}$$

① 概率论的基础定律之一是：对一个任意事件 D 和 t 个相互排斥的事件 $C_1,C_2,\cdots,C_t$，D 的总概率是 $P(D)=P(D\mid C_1)P(C_1)+\cdots+P(D\mid C_t)P(C_t)$。

则全部输出字母的概率分布可以根据下式计算：

$$\mathbf{v} = \mathbf{Qz} \tag{8.3.6}$$

具有元素 $q_{kj} = P(b_k \mid a_j)$ 的矩阵 **Q** 称为正向信道传输矩阵或简写为信道矩阵。

为了决定一个带有正向信道传输矩阵 **Q** 的信道的容量，必须在信息用户观察某一特定的输出 b_k 的假设下，首先计算信息源的熵。式(8.3.4)定义了对任意观察的 b_k 的一个信源符号的分布，所以每个 b_k 有一个条件熵函数，根据推导式(8.3.3)的步骤，这个条件熵函数用 $H(\mathbf{z} \mid b_k)$ 表示，可以写为：

$$H(\mathbf{z} \mid b_k) = -\sum_{j=1}^{J} P(a_j \mid b_k)\log P(a_j \mid b_k) \tag{8.3.7}$$

这里 $P(a_j \mid b_k)$ 是在用户接收到 b_k 的条件下，符号 a_j 由信源传送的概率。这个表达式对所有 b_k 的期望值为：

$$H(\mathbf{z} \mid \mathbf{v}) = \sum_{k=1}^{K} H(\mathbf{z} \mid b_k)P(b_k) \tag{8.3.8}$$

在将 $H(\mathbf{z} \mid b_k)$ 用式(8.3.7)进行替换并且进行某些小的调整① 后，公式可以写成：

$$H(\mathbf{z} \mid \mathbf{v}) = -\sum_{j=1}^{J}\sum_{k=1}^{K} P(a_j, b_k)\log P(a_j \mid b_k) \tag{8.3.9}$$

这里 $P(a_j, b_k)$ 是 a_j 和 b_k 的联合概率。就是说，$P(a_j, b_k)$ 是传送 a_j 并同时接收 b_k 的概率。

$H(\mathbf{z} \mid \mathbf{v})$ 项被称为 **z** 关于 **v** 的条件信息量总平均值(疑义度)。它表示观察由它引起的输出符号的假设下信源符号的平均信息。由于 $H(\mathbf{z})$ 是一个信源符号的平均信息，假设不知道输出符号，$H(\mathbf{z})$ 和 $H(\mathbf{z} \mid \mathbf{v})$ 之间的差别是在观察一个单一输出符号时接收到的平均信息。这种差别表示为 $I(\mathbf{z}, \mathbf{v})$，并被称为 **z** 和 **v** 的互信息，定义为：

$$I(\mathbf{z}, \mathbf{v}) = H(\mathbf{z}) - H(\mathbf{z} \mid \mathbf{v}) \tag{8.3.10}$$

用式(8.3.3)和式(8.3.9)代替 $H(\mathbf{z})$ 和 $H(\mathbf{z} \mid \mathbf{v})$，并且使用 $P(a_j) = P(a_j, b_1) + P(a_j, b_2) + \cdots + P(a_j, b_K)$，得到：

$$I(\mathbf{z}, \mathbf{v}) = \sum_{j=1}^{J}\sum_{k=1}^{K} P(a_j, b_k)\log \frac{P(a_j, b_k)}{P(a_j)P(b_k)} \tag{8.3.11}$$

这个公式在经过进一步的处理后，可以写成：

$$I(\mathbf{z}, \mathbf{v}) = \sum_{j=1}^{J}\sum_{k=1}^{K} P(a_j) q_{kj} \log \frac{q_{kj}}{\sum_{i=1}^{J} P(a_i) q_{ki}} \tag{8.3.12}$$

这样，观察信道的单一输出时接收到的平均信息是输入或信源符号概率向量 **z** 和信道矩阵 **Q** 的函数。$I(\mathbf{z}, \mathbf{v})$ 的最小可能值为零并且在输入和输出符号统计上独立的时候发生，此时，$P(a_j, b_k) = P(a_j)P(b_k)$，且对所有的 j 和 k，式(8.3.11)的对数项为 0。在向量 **z** 中所有信源概率的可能选择中，$I(\mathbf{z}, \mathbf{v})$ 的最大值就是由信道矩阵 **Q** 描绘的信道容量 C。即，

$$C = \max_{\mathbf{z}}[I(\mathbf{z}, \mathbf{v})] \tag{8.3.13}$$

这里最大值取自所有可能的输入符号的概率。信道的容量定义了能够通过信道可靠地传送信息的最大传送率(每个信源符号 m 元信息单元)。此外，信道的容量并不取决于信源的输入概

---

① 这种用法解释为，两个事件 C 和 D 的联合概率是 $P(C, D) = P(C)P(D \mid C) = P(D)P(C \mid D)$。

率(即,取决于如何使用信道),而是独立定义信道的条件概率函数。

例 8.6　二值情况

考虑一个带有信源字母表 $A=\{a_1,a_2\}=\{0,1\}$ 的二值信息源。信源生成符号 a_1 和 a_2 的概率分别是 $P(a_1)=p_{bs}$, $P(a_2)=1-p_{bs}=\bar{p}_{bs}$。根据式(8.3.3),信源的熵是:

$$H(\mathbf{z}) = -p_{bs}\log_2 p_{bs} - \bar{p}_{bs}\log_2 \bar{p}_{bs}$$

因为 $\mathbf{z}=[P(a_1),P(a_2)]^T=[p_{bs},1-p_{bs}]^T$, $H(\mathbf{z})$取决于单一的参量 p_{bs},且等式的右边被称为二值熵函数,表示为 $H_{bs}(\cdot)$。比如 $H_{bs}(t)$是函数 $-t\log_2 t-\bar{t}\log_2 \bar{t}$。图 8.8(a)显示了当 $0\leqslant p_{bs}\leqslant 1$ 时 $H_{bs}(p_{bs})$的曲线。注意,H_{bs}当 p_{bs}为 1/2 时得到最大值(1 比特)。对所有其他的 p_{bs}值,信源提供小于 1 比特的信息。

现在假设通过一个带有噪声的二值信道传送信息,并令在传送任何符号过程中出现一个错误的概率为 p_e。这样一个信道被称为一个二值对称信道(BSC),并用信道矩阵定义为:

$$\mathbf{Q} = \begin{bmatrix} 1-p_e & p_e \\ p_e & 1-p_e \end{bmatrix} = \begin{bmatrix} \bar{p}_e & p_e \\ p_e & \bar{p}_e \end{bmatrix}$$

对于每个输入或信源符号,BSC 根据输出字母表 $B=\{b_1,b_2\}=\{0,1\}$生成一个输出 b_j。接受输出符号 b_1 和 b_2 的概率可以由式(8.3.6)决定:

$$\mathbf{v} = \mathbf{Q}\mathbf{z} = \begin{bmatrix} \bar{p}_e & p_e \\ p_e & \bar{p}_e \end{bmatrix}\begin{bmatrix} p_{bs} \\ \bar{p}_{bs} \end{bmatrix} = \begin{bmatrix} \bar{p}_e p_{bs} + p_e \bar{p}_{bs} \\ p_e p_{bs} + \bar{p}_e \bar{p}_{bs} \end{bmatrix}$$

因此,由于 $\mathbf{v}=[P(b_1),P(b_2)]^T=[P(0),P(1)]^T$,输出为 0 的概率是 $\bar{p}_e p_{bs}+p_e\bar{p}_{bs}$,输出为 1 的概率是 $p_e p_{bs}+\bar{p}_e\bar{p}_{bs}$。

BSC 的互信息可以由式(8.3.12)进行计算。展开这个公式的和式并对适当的项进行整理后得到:

$$I(\mathbf{z},\mathbf{v}) = H_{bs}(p_{bs}p_e + \bar{p}_{bs}\bar{p}_e) - H_{bs}(p_e)$$

这里 $H_{bs}(\cdot)$是图 8.8(a)的二值熵函数。对 p_e 的一个固定值,当 p_{bs}为 0 或 1 时,$I(\mathbf{z},\mathbf{v})$是 0。另外,当二值信源符号等概率出现时,$I(\mathbf{z},\mathbf{v})$取其最大值。图 8.8(b)显示了对于所有 p_{bs}值和一个给定的信道误差 p_e 的 $I(\mathbf{z},\mathbf{v})$。

根据式(8.3.13),由以所有可能的信源分布为基础的互信息的最大值得到 BSC 容量。根据图 8.8(b),对于所有可能的二值信源分布(即,对于 $0\leqslant p_{bs}\leqslant 1$ 或对于 $\mathbf{z}=[0,1]^T$ 到 $\mathbf{z}=[1,0]^T$)的曲线 $I(\mathbf{z},\mathbf{v})$,可以看到当 $p_{bs}=1/2$ 时,$I(\mathbf{z},\mathbf{v})$取最大值(对任意的 p_e)。p_{bs}的值对应于信源概率向量 $\mathbf{z}=[1/2,1/2]^T$。相应的 $I(\mathbf{z},\mathbf{v})$值为 $1-H_{bs}(p_e)$。因此 BSC 的容量(图 8.8(c)中的曲线)是:

$$C = 1 - H_{bs}(p_e)$$

注意,当信道不可能出错($p_e=0$)或必然出错($p_e=1$)时,信道的容量取得 1 比特/符号的最大值。无论是哪种情况,都可以传送最大信息,因为信道的输出是完全可预测的。然而,当 $p_e=1/2$ 时,信道的输出是完全不可预测的并且不能通过这个信道传送任何信息。

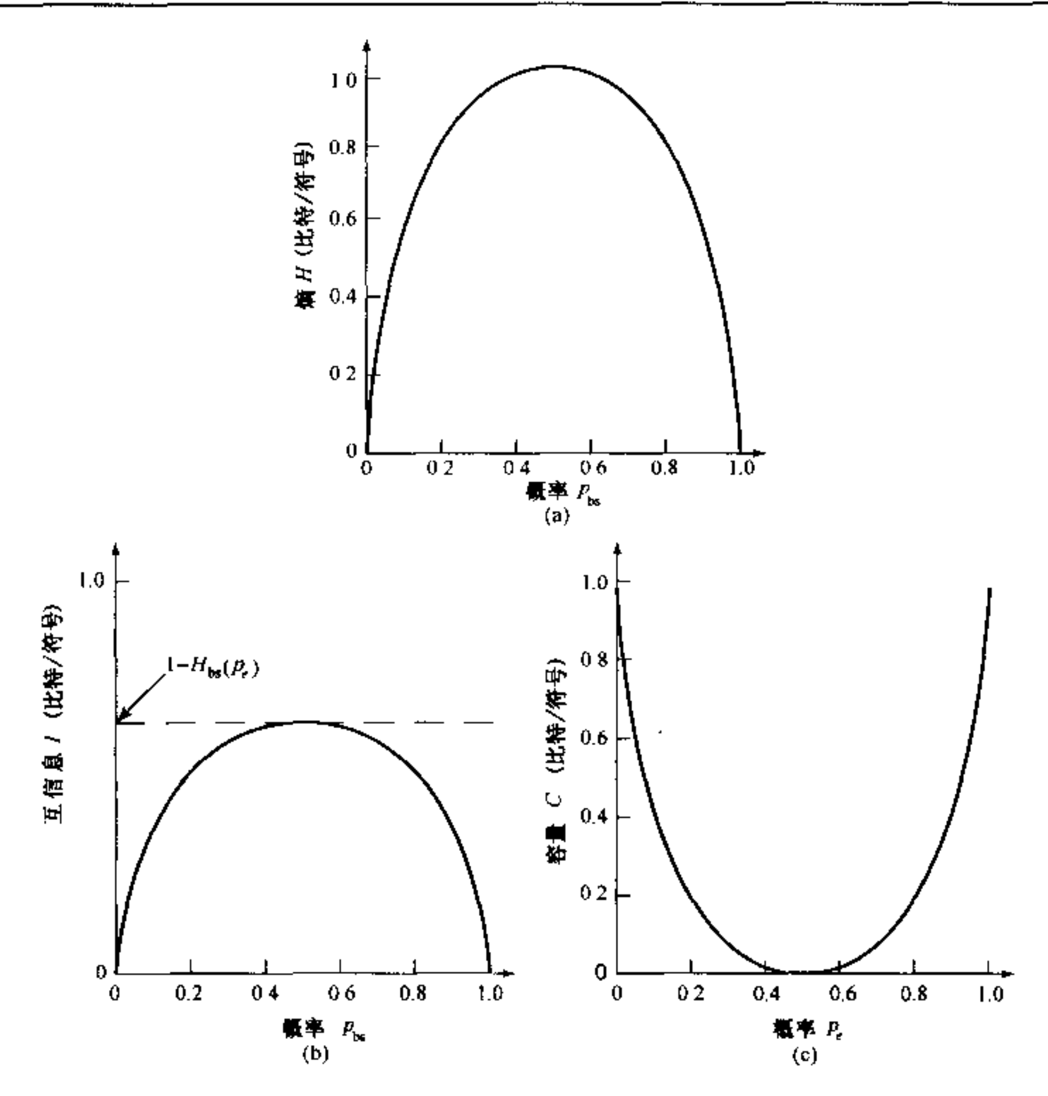


图 8.8 三个二值信息函数。(a)二值熵函数,(b)二值对称信道(BSC)的互信息,(c)BSC的容量

8.3.3 基本编码定理

8.3.2 节中介绍的总体数学框架是以图 8.7 中显示的模型为基础的。图 8.7 中包含一个信源、信道和信宿。这一节,在模型中增加了一个通信系统,并分析关于编码或信息表示的三个基本理论。如图 8.9 所示,通信系统插在信源和信宿之间,它包含一个编码器和一个解码器。

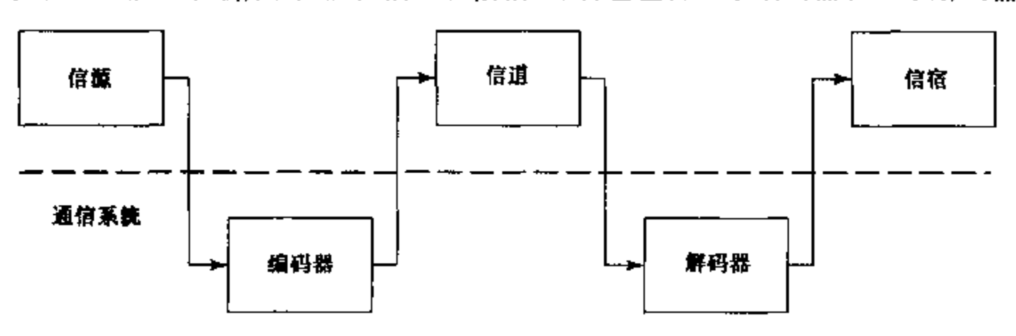


图 8.9 一个通信系统的模型

无噪声编码定理

当信道和通信系统中不存在噪声的时候，通信系统的主要功能是用尽可能简捷的方法表示信源。在这样的情况下，无噪声编码定理，又称香农第一定理（香农，Shannon[1948]），定义了可以达到的每个信源符号的最小平均码字长度。

一个具有有限集合$(A,\mathbf{z})$和在统计上独立的符号源的信息源被称为零记忆信源。如果将它的输出看做来自信源字母表的（而不是一个单一的符号）符号的一个 n 元组，信源输出就取自所有可能的 n 个元素序列集合$A'=\{\alpha_1,\alpha_2,\cdots,\alpha_{J^n}\}$的 J^n 个可能值中的一个，表示为 α_i。换句话说，每个 α_i（称为一个块随机变量）由来自 A 的 n 个符号组成（符号 A'区别于来自 A 的块符号集合，A 是单个符号集合）。一个给定的 α_i 的概率是 $P(\alpha_i)$，…这个概率是与单符号概率 $P(a_j)$通过下式相联系的：

$$P(\alpha_i)=P(a_{j1})P(a_{j2})\cdots P(a_{jn}) \tag{8.3.14}$$

这里，下标 $j1,j2,\cdots,jn$ 用于指示来自集合A 的n 个符号，它构成了一个 α_i。如前所述，向量 $\mathbf{z}'$（撇号表示使用了块随机变量）表示所有信源概率的集合$\{P(\alpha_1),P(\alpha_2),\cdots,P(\alpha_{J^n})\}$，信源的熵为：

$$H(\mathbf{z}')=-\sum_{i=1}^{J^n}P(\alpha_i)\log P(\alpha_i)$$

用式(8.3.14)代替 $P(\alpha_i)$经过化简得到：

$$H(\mathbf{z}')=nH(\mathbf{z}) \tag{8.3.15}$$

因此，零记忆信息源（这个信源产生块随机变量）的熵为对应的单符号信源熵的 n 倍。这样一个信源被称为单一符号或非扩充信源的第 n 次扩充。注意信源的首次扩充都是非扩充信源本身。

因为信源输出 α_i 的自信息是 $\log[1/P(\alpha_i)]$，所以用一个整数长度 $l(\alpha_i)$的码字对 α_i 进行编码似乎是合理的。即：

$$\log\frac{1}{P(\alpha_i)}\leqslant l(\alpha_i)<\log\frac{1}{P(\alpha_i)}+1 \tag{8.3.16}$$

在直观上的提示：信源输出 α_i 是用一个码字表示的，这个码字的长度为超出 α_i 自信息的最小整数。[①] 用 $P(\alpha_i)$与这个结果相乘并将所有乘积相加得到：

$$\sum_{i=1}^{J^n}P(\alpha_i)\log\frac{1}{P(\alpha_i)}\leqslant\sum_{i=1}^{J^n}P(\alpha_i)l(\alpha_i)<\sum_{i=1}^{J^n}P(\alpha_i)\log\frac{1}{P(\alpha_i)}+1$$

或

$$H(\mathbf{z}')\leqslant L'_{avg}<H(\mathbf{z}')+1 \tag{8.3.17}$$

这里 L'_{avg}表示对应于非扩充倍源的第 n 次扩充的编码平均字长。即，

$$L'_{avg}=\sum_{i=1}^{J^n}P(\alpha_i)l(\alpha_i) \tag{8.3.18}$$

用 n 除式(8.3.17)并注意到式(8.3.15)中的 $H(\mathbf{z}')/n$ 等于 $H(\mathbf{z})$，得到：

---

① 在这种限制之下可以构造一个惟一的可解的码。

$$H(\mathbf{z}) \leqslant \frac{L'_{avg}}{n} < H(\mathbf{z}) + \frac{1}{n} \tag{8.3.19}$$

在取极限情况下,变成:

$$\lim_{n \to \infty}\left[\frac{L'_{avg}}{n}\right] = H(\mathbf{z}) \tag{8.3.20}$$

式(8.3.19)阐述了一个零记忆信源的香农第一定理。这个定理说明通过对无限扩充的信源进行编码可以令 L'_{avg}/n 无限地接近 $H(\mathbf{z})$。尽管这个结论是在假设信源符号在统计上相互独立的情况下推导出来的,但这个结论可以很容易推广到更一般的信源,这些信源的符号 a_j 的出现取决于位于它前面的有限个符号。这些类型的信源(称为马尔可夫信源)通常用于图像中像素间相关的模型。因为 $H(\mathbf{z})$是 L'_{avg}/n 的下限[即,随着 n 的增大,式(8.3.20)中 L'_{avg}/n 的极限为$H(\mathbf{z})$],所以任何编码策略的效率 η 可以定义为:

$$\eta = n\frac{H(\mathbf{z})}{L'_{avg}} \tag{8.3.21}$$

例 8.7 扩充编码

一个带有信源字母表 $A = \{a_1, a_2\}$的零记忆信息源具有符号概率 $P(a_1) = 2/3$ 和 $P(a_2) = 1/3$。根据式(8.3.3),这个信源的熵为 0.918 比特/符号。如果符号 a_1 和 a_2 由二进制码字 0 和 1 表示,则 $L'_{avg} = 1$ 比特/符号,并且编码的效率为 $\eta = (1)(0.918)/1$ 或 0.918。

表 8.4 总结了刚才提到的编码及一个基于信源第二次扩充的可作为替代的编码过程。表 8.4的下半部分列出了信源的第二次扩充中的四个块符号分别为 $\alpha_1, \alpha_2, \alpha_3$ 和 α_4。根据式(8.3.14),它们的概率分别为 4/9,2/9,2/9 和 1/9。根据式(8.3.18),第二次编码的平均码长是 17/9 或 1.89 比特/符号。第二次扩充的熵是非扩充源的熵的两倍,或者 1.83 比特/符号,因此,第二次编码的效率为 $\eta = 1.83/1.89 = 0.97$。这比非扩充编码的效率0.92稍好。对信源的第二次扩充编码将每个信源符号的编码比特平均数从 1 比特/符号减少到 1.89/2 或 0.94 比特/符号。

噪声编码定理

如果图 8.9 的信道是有噪声的或容易出错的,则关注的重点从以尽可能紧凑的信息表示的编码转到尽可能稳定的通信。这样很自然地会提出这样的问题:在通信过程中到底允许出现多小的错误呢?

例 8.8 噪声二值信道

假设 BSC 有出错概率 $p_e = 0.01$(即,99% 的信源符号可以通过信道正确传送)。一个简单的增强通信可靠性的方法是将每个信息或二进制符号重复几遍。假设传送 0 和 1,编码信息可以使用 000 和 111。在传送三符号信息的过程中无错误发生的概率或者是 $(1 - p_e)^3$或者是 $\bar{p}_e^3$。单个错误发生的概率是 $3p_e\bar{p}_e^2$。两个错误的概率是 $3p_e^2\bar{p}_e$,三个错误的概率是 p_e^3。因为单一符号传通错误发生的概率小于 50%,可以通过三个接收符号进行多数投票的方法对接收信息进行解码。因此,对三符号码字进行错误解码的概率是两个符号错误和三个符号错误的概率之和,即 $p_e^3 + 3p_e^2\bar{p}_e$。当无错误或单一错误发生时,多数投票可以将信息正确解码。对 $p_e = 0.01$,出现一个通信错误的概率是 0.0003。

表 8.4 扩充编码示例

α_i	信源符号	$P(\alpha_i)$,式(8.3.14)	$I(\alpha_i)$,式(8.3.1)	$l(\alpha_i)$,式(8.3.16)	码字	码长
第一次扩充						
α_1	a_1	2/3	0.59	1	0	1
α_2	a_2	1/3	1.58	2	1	1
第二次扩充						
α_1	a_1a_1	4/9	1.17	2	0	1
α_2	a_1a_2	2/9	2.17	3	10	2
α_3	a_2a_1	2/9	2.17	3	110	3
α_4	a_2a_2	1/9	3.17	4	111	3

通过对刚才描述的重复编码方案的扩充,将通信中发生的错误控制在希望的尽可能小的程度内。在一般情况下,通过使用长度为 r 的 K 元编码序列对信源进行第 n 次扩充来编码,这里 $K' \leqslant J^n$。这种方法的关键是,在 K' 个可能的编码序列中选择合适的 φ 作为有效的码字并建立一条规则使正确解码的概率得到最优化。在前面的例子中,将每个信源符号重复三次等同于,使用 8 个之中的两个可能出现的二进制码字,对非扩充二值信源进行块编码。两个有效的码字是 000 和 111。如果一个无效码字输送到解码器中,则由三个编码比特的多数投票决定输出。

一个零记忆信息源以一定速率(以每个符号的信息单元数表示)生成信息,速率等于这个信源的熵 $H(\mathbf{z})$。信源的第 n 次扩充以每个符号 $H(\mathbf{z}')/n$ 信息单元的速度提供信息。如果信息是进行过编码的,就像前面的例子那样,则当用于对信源编码的 φ 个有效码字等概率出现时,编码信息的最大速率是 $\log(\varphi/r)$。因此,尺寸为 φ 且块长度为 r 的编码具有:

$$R = \log\frac{\varphi}{r} \quad \text{信息单元 / 符号} \tag{8.3.22}$$

的速率。香农第二定理(Shannon[1948])也称为噪声编码定理,告诉我们对任意 $R < C$,存在一个整数 r 和块长度为 r、速率为 R 的编码,这个编码的块解码误差的概率小于或等于任意的 $\varepsilon > 0$,这里 C 是具有矩阵 $\mathbf{Q}$ 的零记忆信道[①] 的容量。因此,只要编码信息率小于信道容量,则出现误差的概率就可以为任意小。

信源编码定理

这个定理描述了迄今为止在可靠信道和不可靠信道上进行无误差通信时建立的基本限制。在这一节中,转向信道无误差而通信过程本身有损耗的情况上来。在这样的环境下,通信系统的主要功能是"信息压缩"。在大多数情况下,由压缩引入的平均差错受到某个最大允许级别 D 的限制。我们希望判定在给定的保真度准则限制下,信源向信宿传送信息的最小速率。这个问题在称为"率失真理论"的信息论分支中进行了明确的讲述。

分别使用有限集合$(A,\mathbf{z})$和$(B,\mathbf{z})$定义图 8.9 中的信源和解码输出。假设图 8.9 的信道

① 一个零记忆信道是指信道对当前输入的响应与对前一个输入符号的响应是彼此独立的。

是不出现错误的,因此,根据式(8.3.6)将 $\mathbf{z}$ 和 $\mathbf{v}$ 相联系的信道矩阵 $\mathbf{Q}$ 可以被认为是将编解码处理进行模型化的过程。因为编解码处理是确定性的,$\mathbf{Q}$ 描述了一个人工零记忆信道,它模型化了信息压缩和解压缩的功能。每次信源生成信源符号 a_j 时,都用一个编码符号表示,这个编码符号以概率 q_{kj} 解码为输出符号 b_k(见8.3.2节)。

解决信源编码问题,使其平均失真小于 D,需要对每个可能的信源输出的近似值指定一个数量的失真值。对非扩充信源的简单情况,一个称为失真度量的非负值开销函数 $\rho(a_j, b_k)$ 可用于定义再次生成解码输出 b_k 的信源输出 a_j 时进行的相应处理。信源的输出是随机的,因此,其失真同样是随机变量,这个随机变量用 $d(\mathbf{Q})$ 表示为:

$$\begin{aligned} d(\mathbf{Q}) &= \sum_{j=1}^{J}\sum_{k=1}^{K}\rho(a_j, b_k)P(a_j, b_k) \\ &= \sum_{j=1}^{J}\sum_{k=1}^{K}\rho(a_j, b_k)P(a_j)q_{kj} \end{aligned} \tag{8.3.23}$$

符号 $d(\mathbf{Q})$ 强调平均失真是编解码处理过程的一个函数。编解码过程用 $\mathbf{Q}$(前面提到过的)来模型化。当且仅当与 $\mathbf{Q}$ 对应的平均失真小于或等于 D 时,特定的编解码过程才被称为是 D 所接纳的。因此,所有 D 所接纳的编解码过程的集合表示为:

$$\mathbf{Q}_D = \{q_{kj} \mid d(\mathbf{Q}) \leqslant D\} \tag{8.3.24}$$

因为每个编解码过程都是用一个人工信道矩阵 $\mathbf{Q}$ 定义的,所以根据观察单一解码器的输出得到的平均信息可以根据式(8.3.12)进行计算。因此,可以定义一个率失真函数:

$$R(D) = \min_{\mathbf{Q}\in\mathbf{Q}_D}[I(\mathbf{z}, \mathbf{v})] \tag{8.3.25}$$

这里,假设式(8.3.12)的最小值超出了 D 所接纳的编码。注意,最小值可以在 $\mathbf{Q}$ 上取得,因为 $I(\mathbf{z}, \mathbf{v})$ 是向量 $\mathbf{z}$ 中的概率和矩阵 $\mathbf{Q}$ 中元素的函数。如果 $D=0$,则 $R(D)$ 小于或等于信源的熵,或 $R(0) \leqslant H(\mathbf{z})$。

式(8.3.25)定义了在限制平均失真小于或等于 D 的条件下,信源可以传送给信宿的信息的最小速率。为了计算这个速率[即,$R(D)$],可以简单地通过对受下列限制的 $\mathbf{Q}$(或 q_{kj})进行适当的选择最小化 $I(\mathbf{z}, \mathbf{v})$。

$$q_{kj} \geqslant 0 \tag{8.3.26}$$

$$\sum_{k=1}^{K} q_{kj} = 1 \tag{8.3.27}$$

和

$$d(\mathbf{Q}) = D \tag{8.3.28}$$

式(8.3.26)和式(8.3.27)是信道矩阵 $\mathbf{Q}$ 的基本性质。$\mathbf{Q}$ 的元素必须为正,且因为对于任何产生的输入符号有些输出是必然会收到的,所以 $\mathbf{Q}$ 中每一列的和必须为1。式(8.3.28)指出当允许出现最大可能失真的时候,就会得到最小信息速率。

例8.9 计算一个零记忆二值信源的率失真函数

考虑一个具有等概率信源符号{0,1}的零记忆二值信源,其简单失真量度为:

$$\rho(a_j, b_k) = 1 - \delta_{jk}$$

这里 δ_{jk} 是单位德尔塔函数。因为如果 $a_j \neq b_k$,则 $\rho(a_j, b_k)$ 等于1,否则为0,所以每个编

码解码误差被记为一个单位的失真。这个变量的微分可以用于计算 $R(D)$。令 $\mu_1, \mu_2, \cdots, \mu_{J+1}$ 为拉格朗日乘数,构造增强的准则函数:

$$J(\mathbf{Q}) = I(\mathbf{z}, \mathbf{v}) - \sum_{j=1}^{J} \mu_j \sum_{k=1}^{K} q_{kj} - \mu_{J+1} d(\mathbf{Q})$$

令它的 JK 关于 q_{kj} 的导数为 0(即,$\mathrm{d}J/\mathrm{d}q_{kj} = 0$),并连同与式(8.3.27)和式(8.3.28)相关的 $J+1$ 个方程式对未知量 q_{kj} 和 $\mu_1, \mu_2, \cdots, \mu_{J+1}$ 求解,得出方程式。如果得到的 q_{kj} 是非负值[或满足式(8.3.26)],就找到了一个有效的解。对于上述定义的信源和失真对,得到下列 7 个公式(带有 7 个未知量):

$$\begin{aligned}
&2q_{11} = (q_{11} + q_{12})\exp[2\mu_1] && 2q_{22} = (q_{21} + q_{22})\exp[2\mu_2] \\
&2q_{12} = (q_{11} + q_{12})\exp[2\mu_1 + \mu_3] && 2q_{21} = (q_{21} + q_{22})\exp[2\mu_2 + \mu_3] \\
&q_{11} + q_{21} = 1 && q_{12} + q_{22} = 1 \\
&q_{21} + q_{12} = 2D &&
\end{aligned}$$

由此生成了一系列冗长的但形式简单的代数运算步骤:

$$\begin{aligned}
q_{12} &= q_{21} = D \\
q_{11} &= q_{22} = 1 - D \\
\mu_1 &= \mu_2 = \log\sqrt{2(1-D)} \\
\mu_3 &= \log\frac{D}{1-D}
\end{aligned}$$

这样得到:

$$\mathbf{Q} = \begin{bmatrix} 1-D & D \\ D & 1-D \end{bmatrix}$$

由于给出的信源符号是等概率的,所以最大可能失真是 1/2。因此 $0 \leqslant D \leqslant 1/2$ 和 $\mathbf{Q}$ 的元素对所有的 D 都满足式(8.3.12)。与 $\mathbf{Q}$ 相关的互信息和前面定义的二值信源是用式(8.3.12)计算的。然而,注意到在 $\mathbf{Q}$ 和二值对称信道矩阵之间的相似性,可以很快写出:

$$I(\mathbf{z}, \mathbf{v}) = 1 - H_{\mathrm{bs}}(D)$$

这个结果遵循例 8.6,将 $p_{\mathrm{bs}} = 1/2$ 和 $p_e = D$ 代入 $I(\mathbf{z}, \mathbf{v}) = H_{\mathrm{bs}}(p_{\mathrm{bs}} p_e + \bar{p}_{\mathrm{bs}} \bar{p}_e) - H_{\mathrm{bs}}(p_e)$。根据式(8.3.25)率失真函数为:

$$R(D) = \min_{\mathbf{Q} \in \mathbf{Q}_D} [1 - H_{\mathrm{bs}}(D)] = 1 - H_{\mathrm{bs}}(D)$$

最终的简化是基于以下条件,即对给定的 D,$1 - H_{\mathrm{bs}}(D)$ 假定为单值,默认情况下为最小值。得到的函数即图 8.10 中所示的曲线。这条曲线的形状是典型的最大率失真函数。注意,D 的最大值表示为 $D_{\max}$。这个值对于所有的 $D \geqslant D_{\max}$ 有 $R(D) = 0$。另外,$R(D)$ 总是正值,且单调递减,在区间 $(0, D_{\max})$ 内为凸形的。

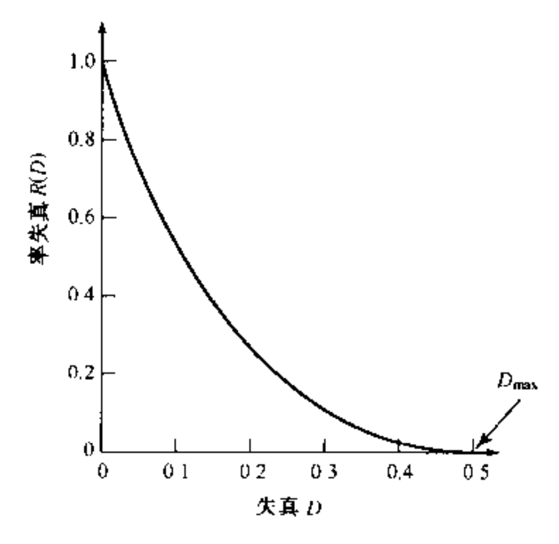


图 8.10 二值对称信源的率失真函数

正如前面的例子,对简单信源和失真量度可计算率失真函数。再有,在分析方法失败或方法实现起来不现实的时候,可以使用适于在数字计算机上实现的收敛的迭代算法。计算出 $R(D)$之后(对任何零记忆信源和单字符失真量度①),信源编码定理告诉我们对任意的 $\varepsilon>0$,存在一个 r 和块长度为 r、传送速率 $R<R(D)+\varepsilon$ 的编码,使得每个字符的平均失真满足 $d(\mathbf{Q})\leqslant D+\varepsilon$的条件。这个定理和噪声编码定理的一个重要的实用结果是,在容量为 $C>R(D)+\varepsilon$的信道所提供的误差概率为任意小的解码器中,可以对信源输出进行复原。这个结果就是众所周知的信息传输定理。

8.3.4 信息论的应用

信息论提供了处理信息表达和直接定量操作的基本工具。在本小节中,将探寻这些工具在特定的图像压缩问题中的应用。因为信息论的基本前提是生成的信息可以被模型化为一个概率过程,所以首先建立图像生成过程的统计模型。

例 8.10 计算图像的熵

考虑简单的 8 比特图像信息内容(或熵)的估计问题:

21	21	21	95	169	243	243	243
21	21	21	95	169	243	243	243
21	21	21	95	169	243	243	243
21	21	21	95	169	243	243	243

一种相对简单的方法是假设一个特定的信源模型,并计算基于这个模型的图像的熵。例如,可以假设图像由虚构的"8 比特灰度级信源"生成,这个信源根据事先定义的概率定律产生在统计上相互独立的像素。在这种情况下,信源符号是有灰度级的,并且信源字符表由 256 个可能的符号构成。如果符号出现的概率是已知的,图像中每个像素的平均信息量或熵可以使用式(8.3.3)计算得到。例如,在具有均匀概率密度的情况下,信源符号是

① 单字符失真量度是指与一个字符(或符号)块中每个字符(或符号)的失真之和相对应的失真。

等概率的，且信源由一个8比特/像素的熵进行表征。即，每个信源输出(像素)的平均信息是8比特。这样一来，前述4×8大小的图像的总平均信息量是256比特。但这幅特殊图像仅仅是由这个信源产生的$2^{8\times4\times8}$或2^{256}($\sim10^{77}$)种等概率的4×8大小图像的一种。

一种可作为替代的估计信息量的方法是，构造一个基于图像中灰度级出现频率的信源模型。即，被观察的图像可以简单解释为生成这幅图像的灰度级信源的特性。因为被观察图像是信源特性惟一有效的指示器，用取样图像的灰度级直方图对信源符号的概率模型化是较为合理的。

灰 度 级	计 数	概 率
21	12	3/8
95	4	1/8
169	4	1/8
243	12	3/8

称为一阶估计的信源熵的估计可以用式(8.3.3)计算。本例子的一阶估计为1.81比特/像素。信源和/或图像的熵就是1.81比特/像素，或总共58比特。

产生取样图像的灰度级信源熵的更好估计，可以通过分析取样图像中像素块的相对频率来计算，这里，一个块就是邻接的像素组。当块的大小接近无限大时，这个估计也接近信源真实的熵(这个结果可以用8.3.3节中证明无噪声编码定理的有效性时所使用的过程进行说明)。因此，通过假设取样图像是线与线首尾相接的，能够计算像素对的相对频率(即，信源的第二次扩充)：

灰度级对	计 数	概 率
(21,21)	8	1/4
(21,95)	4	1/8
(95,169)	4	1/8
(169,243)	4	1/8
(243,243)	8	1/4
(243,21)	4	1/8

得到的熵估计[再次使用式(8.3.3)]是2.5/2，或1.25比特/像素。这里用2除的原因是由于一次同时考虑了两个像素。这个估计被称为信源熵的二阶估计，因为这个估计是通过计算两个像素块的相对频率得到的。尽管三阶、四阶和更高阶的估计可以提供信源熵的更好近似，但它们收敛于真实信源熵的速度太慢并且陷于大量计算之中。例如，普通的8比特图像具有$(2^8)^2$或65 536种可能的符号对，这些符号对的相对频率都必须进行计算。如果考虑一个5像素的块，可能的5元组的数目是$(2^8)^5$种，或10^{12}种。

尽管计算一幅图像实际的熵是困难的，但像前述例子那样对图像进行的估计，提供了一种对图像压缩能力的认识。例如，熵的一阶估计是通过变长编码可以实现压缩的下界(在8.1.1节，变长编码用于减少编码冗余)。另外，熵的高阶估计和一阶估计之间的差异表明了是否存在像素间冗余。即，它们显示出图像中的像素在统计上是否是彼此独立的。如果像素

在统计上是独立的(即,不存在像素间冗余),则高阶估计等于一阶估计,且变长编码提供了最佳的压缩。对于前述例子中考虑的图像,在一阶和二阶估计之间用数字表示的差异表明,可以建立一种映射,以便允许从图像表达中去掉额外的 1.81 - 1.25 = 0.56 比特/像素。

例 8.11 使用映射减少熵

考虑前述例子的图像中像素的映射,构造一种表达方法:

$$\begin{matrix} 21 & 0 & 0 & 74 & 74 & 74 & 0 & 0 \\ 21 & 0 & 0 & 74 & 74 & 74 & 0 & 0 \\ 21 & 0 & 0 & 74 & 74 & 74 & 0 & 0 \\ 21 & 0 & 0 & 74 & 74 & 74 & 0 & 0 \end{matrix}$$

这里,通过复制原图像的第一列并使用相邻列的余下元素的算术差构造一个差异阵列。例如,在新的表示方法中,第 1 行第 2 列的元素是(21 - 21),或 0。得到的差异分布是:

灰度级或差异	计 数	概 率
0	12	1/2
21	4	1/8
74	12	3/8

如果现在考虑"差异信源"生成的映射阵列,可以再次使用式(8.3.3)计算阵列熵的一阶估计,值为 1.41 比特/像素。因此通过对映射差异图像进行变长编码,原图像可以用 1.41 比特/像素或总共 46 比特表示。这个值比前述例子中计算的 1.25 比特/像素的熵的二阶估计要大,因此,我们知道可以找到更好的映射。

前述例子说明一幅图像熵的一阶估计不一定必然是图像的最小码率。原因是图像中的像素一般在统计上不是彼此独立的。正如 8.2 节中注意到的,一幅图像实际熵的最小化过程被称为信源编码。在无错误的情况下,这个过程包括映射和符号编码两种操作。如果信息损失是允许的,这个过程还包括第三步量化操作。

关于有损图像压缩的更为复杂一些的问题同样可以使用信息论这一工具解决。然而,在这种情况下,主要应用信源编码定理。正如 8.3.3 节中指出的那样,这个定理揭示,任何零记忆信源可以用一个速率 $R < R(D)$ 的码进行编码,以便每个符号失真的平均值小于 D。为了将这个结果正确地应用于有损图像压缩,需要确定适当的信源模型,设计一个有意义的失真度量,并计算率失真函数 $R(D)$。这个处理过程的第一步已经研究过了。第二步可以用来自 8.1.4 节的客观保真度准则方便地解决。最后一步涉及寻找某个矩阵 $\mathbf{Q}$,这个矩阵是由式(8.3.24)到式(8.3.28)的限制条件下使式(8.3.12)最小化的元素所组成的。遗憾的是,这个工作非常困难——在实际感兴趣的所有情况中,只有一小部分得到了解决。其中一个是图像为高斯型随机场并且失真度量是一个加权的平方误差函数的情况。此时,最佳编码器必须将图像扩展为它的卡南-洛耶夫(Karhunen-Loève)分量(见 11.4 节)且用相等的均方误差表示每个分量(Davisson[1972])。

8.4　无误差压缩

在众多的应用中，无误差压缩是仅有的可以接受的数据压缩方法。这类应用中的一种是医疗或商业文件的归档。在这些应用场合，有损压缩通常因为法律原因而被禁止。另一种应用是卫星成像的处理。在这类应用中，收集到的数据的使用和所花的费用都使得人们不希望有任何的数据损失。还有一类应用是数字 X 光照相术，这种应用中信息的丢失会导致损坏诊断的精确性。在这一类情况下，无误差压缩的需要是由预期用户和图像性质所推动的。

在本节中，我们的注意力将放在当前应用中的无误差压缩策略上。这些策略通常提供的压缩率为 2 到 10。另外，它们对二值图像和灰度级图像的适用性是相同的。正如 8.2 节中所指出的那样，无误差压缩技术通常由两种彼此独立的操作组成：(1)为减少像素间冗余，建立一种可替代的图像表达方式；(2)对这种表达方式进行编码以便消除编码冗余。这些步骤与图 8.6讨论的信源编码模型的映射和符号编码操作相对应。

8.4.1　变长编码

无误差图像压缩的最简单方法就是减少仅有的编码冗余。编码冗余通常存在于图像灰度级的自然二进制编码过程中。如 8.1.1 节中提到的，它可以通过对灰度级进行编码使式(8.1.4)最小化得到消除。这样做需要变长编码结构，它可把最短的码字赋予出现概率最大的灰度级。这里，对于构造这样的码字分析几种最佳的和接近最佳的编码技术。这些技术都是使用信息论的语言进行表达的。实际上，信源符号既可能是图像灰度级，也可能是灰度级映射操作的输出(像素差异、行程宽度，等等)。

霍夫曼编码

消除编码冗余的最常用技术要归功于霍夫曼(Huffman[1952])。当对独立信息源的符号进行编码的时候，霍夫曼编码对每个信源符号生成可能的最小数量的编码符号。无噪声编码定理(见 8.3.3 节)对于 n 的一个固定值生成的编码是最佳的，但受限于每次只能对一个信源符号进行编码。

霍夫曼方法的第一步是将需要考虑的符号概率进行排序，并将具有最低概率的符号联结为一个单一的符号，用这个符号在信源化简的下一步中替代联结之前的两个符号。图8.11说明了在二值编码情况下的这种处理过程(也可以构造 K 元的霍夫曼编码)。在最左边，一个假设的信源符号集合和它们的概率从上到下按概率值减少的顺序排列。进行第一次信源化简时，底部的两个概率 0.06 和 0.04 联结成为一个概率值为 0.1 的“复合符号”。这个复合符号和它对应的概率被置于第一次信源化简的列中，以便化简后的信源概率仍旧是从大到小排列。这个过程一直重复持续到信源只有两个符号(在最右边)。

霍夫曼编码过程的第二个步骤是对每个化简后的信源进行编码，从最小的信源开始，一直工作到原始的信源。当然，对于一个两个符号信源的最小长度，二值编码是符号 0 和 1。如图 8.12所示，这些符号被分配给最右边的两个符号(这种分配是任意的；掉转 0 和 1 的顺序并无妨碍)。当简化信源中，合并的两个符号产生的概率为 0.6 的简化信源符号赋值时，用 0 去编码，而后将一个 0 和 1 任意地附加在这两个符号编码上面，用于将两个符号彼此区分开。

原始信源		信源化简			
符号	概率	1	2	3	4
a_2	0.4	0.4	0.4	0.4	0.6
a_6	0.3	0.3	0.3	0.3	0.4
a_1	0.1	0.1	0.2	0.3	
a_4	0.1	0.1	0.1		
a_3	0.06	0.1			
a_5	0.04				

图 8.11　霍夫曼信源化简

原始信源			信源化简							
符号	概率	编码	1		2		3		4	
a_2	0.4	1	0.4	1	0.4	1	0.4	1	0.6	0
a_6	0.3	00	0.3	00	0.3	00	0.3	00	0.4	1
a_1	0.1	011	0.1	011	0.2	010	0.3	01		
a_4	0.1	0100	0.1	0100	0.1	011				
a_3	0.06	01010	0.1	0101						
a_5	0.04	01011								

图 8.12　霍夫曼编码分配过程

然后,对每个化简后的信源重复这种操作直到达到原始信源。图 8.12 最左边显示了最终编码。这个编码的平均长度为:

$$L_{avg} = (0.4)(1) + (0.3)(2) + (0.1)(3) + (0.1)(4) + (0.06)(5) + (0.04)(5)$$
$$= 2.2 \text{ 比特 / 符号}$$

其信源的熵为 2.14 比特/符号。根据式(8.3.21)得到的霍夫曼编码的效率为 0.937。

霍夫曼编码过程对一组符号产生最佳编码,并且概率服从一次只能对一个符号进行编码的限制。在编码都建立了之后,编码和/或解码就简单地以查询表的方式完成了。编码本身是一种瞬时惟一的可解块编码。它之所以称为一种块编码,是因为每个信源符号都映射到一个编码符号的固定序列中。这种编码是瞬时的,因为符号串中的每个码字无须参考后继符号就可以进行解码。它又是惟一可解码的,因为任何符号串只能以一种方式进行解码。因此,任何霍夫曼编码的符号串,可以通过从左到右的方式对串中每个符号进行分析来解码。对于图 8.12中的二值编码,对编码串 010100111100 从左到右的扫描显示,第一个有效码字为 01010,这个编码的符号是 a_3。下一个有效的编码是 011,它所对应的符号为 a_1。以这种方式持续下去得到的完整解码信息是 $a_3 a_1 a_2 a_2 a_6$。

其他接近最佳的变长编码

当要对大量的符号进行编码的时候,构造最佳二值霍夫曼编码不是一件简单的工作。对有 J 个信源符号的一般情况,必须进行 $J-2$ 次的信源化简(见图 8.11)和 $J-2$ 次编码分配(见图 8.12)。因此对具有 256 个灰度级的图像构造最佳霍夫曼编码需要 254 次信源化简和 254 次编码分配。考虑到这项工作在计算上的复杂性,牺牲编码效率以换取编码构造的简单性有时是必要的。

表8.5说明了四种变长编码，它提供了这样一种折中方案。注意，霍夫曼编码的平均长度——表的最后一行——小于列出的其他编码。自然二进制码具有最大的平均长度。另外，霍夫曼技术达到的4.05比特/符号的编码速率接近由式(8.3.3)计算出来的4.0比特/符号的信源熵的底线，这个数字列在表的底部。尽管表8.5中没有其他编码的效率能与霍夫曼编码相比，但所有的编码都易于构造。像霍夫曼编码技术一样，它们都将最短码字分配给最容易出现的信源符号。

表8.5 变长编码

信源符号	概率	二值编码	霍夫曼编码	截取霍夫曼编码	B_2编码	二值移位	霍夫曼移位
块1							
a_1	0.2	00000	10	11	C00	000	10
a_2	0.1	00001	110	011	C01	001	11
a_3	0.1	00010	111	0000	C10	010	110
a_4	0.06	00011	0101	0101	C11	011	100
a_5	0.05	00100	00000	00010	C00C00	100	101
a_6	0.05	00101	00001	00011	C00C01	101	1110
a_7	0.05	00110	00010	00100	C00C10	110	1111
块2							
a_8	0.04	00111	00011	00101	C00C11	111000	0010
a_9	0.04	01000	00110	00110	C01C00	111001	0011
a_{10}	0.04	01001	00111	00111	C01C01	111010	00110
a_{11}	0.04	01010	00100	01000	C01C10	111011	00100
a_{12}	0.03	01011	01001	01001	C01C11	111100	00101
a_{13}	0.03	01100	01110	100000	C10C00	111101	001110
a_{14}	0.03	01101	01111	100001	C10C01	111110	001111
块3							
a_{15}	0.03	01110	01100	100010	C10C10	111111000	000010
a_{16}	0.02	01111	010000	100011	C10C11	111111001	000011
a_{17}	0.02	10000	010001	100100	C11C00	111111010	0000110
a_{18}	0.02	10001	001010	100101	C11C01	111111011	0000100
a_{19}	0.02	10010	001011	100110	C11C10	111111100	0000101
a_{20}	0.02	10011	011010	100111	C11C11	1111111010	0001110
a_{21}	0.01	10100	011011	101000	C00C00C00	111111110	00001111
熵	4.0						
平均码长		5.0	4.05	4.24	4.65	4.59	4.13

表8.5的第5列说明了一个对基本的霍夫曼编码策略的简单修正，它被称为截尾霍夫曼编码。一个截尾霍夫曼编码是通过对信源的具有最大概率的 ψ 个符号进行霍夫曼编码产生的，ψ 是小于 J 的正整数。后面跟有适当的固定长度编码的前缀编码用于表示其他所有的信

源符号。在表 8.5 中,将 ψ 任意地选择为 12,并且生成前缀编码作为第 13 个霍夫曼码字。即,在对 12 个最有可能出现的信源符号进行霍夫曼编码时,将概率为符号 a_{13} 到 a_{21} 的概率之和的一个"前缀符号"包括进来作为第 13 个符号。剩下的 9 个符号的编码用前缀编码进行编码,这些前缀编码变成了 10 和一个等于符号下标减去 13 的 4 位二进制值。

表 8.5 的第 6 列说明了被称为 B 编码的第二种接近最佳的变长编码。当信源符号概率遵从形式如下的幂律的时候,对某些正常数 β 和归一化常数 $c = 1/\sum_{j=0}^{J} j^{-\beta}$,这种编码接近最佳。

$$P(a_j) = cj^{-\beta} \tag{8.4.1}$$

例如,用二值表示法表示的典型打字机打印的文件的行程长度分布是指数的。如表 8.5 所显示的,每个码字都由表示为 $c = 1/\sum_{j=0}^{J} j^{-\beta}$ 的延伸比特和信息比特组成,这些位都是自然二进制数。延伸比特的惟一目的就是将独立的码字分割开,以便对串中的每个码字可以在 0 和 1 之间进行简单的交换。表 8.5 中的 B 编码称为 B_2 编码,因为每个延伸比特使用两个信息比特。对应于信源符号串 $a_{11}a_2a_7$ 的 B_2 编码序列是 001 010 101 000 010 或 101 110 001 100 110,这取决于第一位延伸比特被假设为 0 还是 1。

表 8.5 中剩下的两个变长码称为移位编码。移位编码的生成通过下列步骤:(1)对信源符号进行排列以便符号的概率呈单调递减的顺序;(2)将所有的符号分割为相等大小的符号块;(3)在所有的块中对单个元素进行相同的编码;(4)对每个块增加特定的上移和/或下移符号以便进行块的识别。解码器每次识别一个上移或下移符号,这个符号根据预先定义的基准块上移或下移一个块。

为了生成表 8.5 中第 7 列的 3 位二值移位编码,21 个信源符号根据它们出现概率的不同进行排序,并将它们分成 3 个块,每个块有 7 个符号。上部块的单独符号(a_1 到 a_7)——考虑基准块——用二进制编码 000 到 110 进行编码。第 8 个二进制编码(111)没有被包含在基准块中;相反,这个二进制编码作为一个惟一的上移控制用于识别剩下的块(此时,没有用到下移符号)。剩下的两个块中的符号使用一个或两个上移符号在用于基准块编码的二进制码组合中进行编码。例如,符号 a19 的编码是 111 111 100。

表 8.5 中第 8 列的霍夫曼移位编码是用相似的方式生成的。主要的差别是,霍夫曼编码对基准块进行编码之前要分配移位符号的概率。正常情况下,这种分配是通过将基准块之外的所有信源符号的概率进行相加实现的;即,使用定义截尾霍夫曼编码前缀符号的相同概念。这里,从符号 a_8 到 a_{21} 的概率和是 0.39。移位符号是具有最大概率的符号,且被分配一个最短的霍夫曼码字(00)。

算术编码

与前述的变长编码不同,算术编码生成的是非块码。在算术编码领域中,可以追溯到 Elias(见 Abramson[1963])的研究工作,在信源符号和码字之间是不存在一一对应的关系的。相反,算术编码是给整个信源符号(或消息)序列分配一个单一的算术码字。这个码字本身定义了一个介于 0 和 1 之间的实数间隔。当消息中的符号数目增加时,用于描述消息的间隔变得更小,而表示间隔所需要的信息单元(如,位)的数目变得更多了。

消息的每个符号根据符号出现的概率减小间隔的大小。因为这种技术不像霍夫曼编码方法那样要求将每个信源符号转换成符号的整数(即,每次对一个符号进行编码),所以这种技术达到了(仅在理论上)由8.3.3节中的无噪声编码准则所设定的界限。

图8.13说明了算术编码的基本过程。这里,一个五符号序列(或称消息)为 $a_1a_2a_3a_3a_4$,是来自一个编码的四符号信源。在编码处理的开始,消息被假定为占据整个半开区间[0,1)。如表8.6所示,这个区间开始时根据每个信源符号的出现概率分成四个区域。例如,符号 a_1 对应子区间[0,0.2)。因为 a_1 是消息中进行编码的第一个符号,消息的间隔开始被限制在[0,0.2)的窄区间内。因此,在图8.13中,区间[0,0.2)被扩展到图形的全高度,并且在区间的末端用这个狭窄域的值进行了标注。这个窄域再根据初始信源符号的出现概率进行细分,而后接着处理下一个消息符号。用这种方式,符号 a_2 将子区间变窄为[0.04,0.08),a_3 进一步将子区间变窄为[0.056,0.072),如此进行下去。最后的消息符号必须被保留以作为特定的消息结束指示符,它将子区间变窄为[0.067 52,0.068 8)。当然,在这个子区间内的任何数字(例如,0.068)都可以被用来表示这个消息。

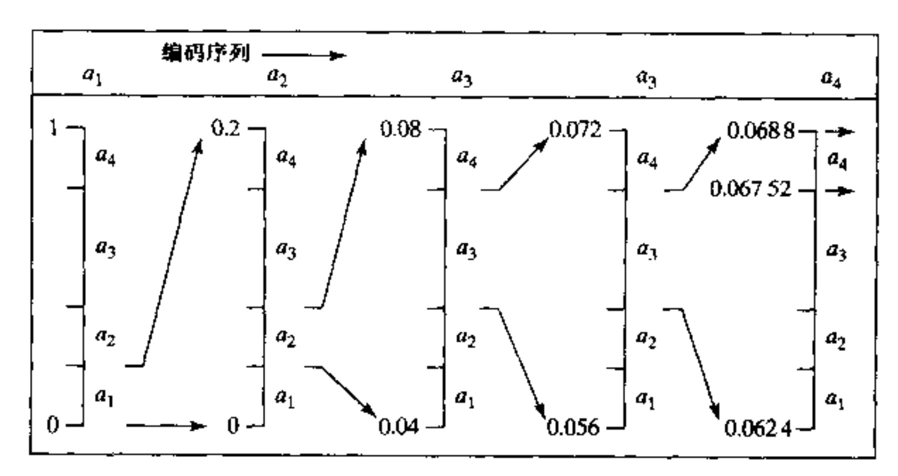


图8.13 算术编码过程

表8.6 算术编码示例

信源符号	概率	初始子区间
a_1	0.2	[0.0,0.2)
a_2	0.2	[0.2,0.4)
a_3	0.4	[0.4,0.8)
a_4	0.2	[0.8,1.0)

在图8.13中的算术编码消息中,使用三个十进制数字表示五符号消息。这样的转换得到每信源符号3/5或0.6个十进制数字,而根据式(8.3.3)与之相比的信源的熵为0.58个十进制数字或10元单元/符号。当编码序列的长度不断增加的时候,得到的算术编码也接近无噪声编码定理所设定的界限。实际上,有两个因素使编码的效能无法达到这个界限:(1)为了将一个消息同其他的消息分离开,增加了消息结束指示符,(2)使用的算法精度是有限的。算术编码的实际实现通过引入尺度策略和舍入策略解决第二个问题(Langdon 和 Risssanen[1981])。尺度策略在根据符号出现概率将子区间进行再次细分之前,将每个子区间重新归一化回[0,1)

区间的范围内。舍入策略保证根据算法的有限精度进行的截尾不会影响根据正确的表示对子区间进行编码。

8.4.2 LZW编码

在分析了消除编码冗余的主要方法之后,现在考虑一种处理图像的像素间冗余的无误差压缩技术。这种技术叫做Lempel-Ziv-Welch(LZW)编码,它对信源符号的可变长度序列分配固定长度码字,且不需要了解有关被编码符号的出现概率的知识。回顾8.3.3节中香农第一定理阐明,一个零记忆信源的第n次扩充可以用比非扩充信源本身更少的表示每个信源符号所需的平均比特数编码。尽管这种压缩技术必须得到美国第4588302号专利的许可,但LZW压缩技术已经被收入主流的图像文件格式中,其他的格式包括图形交换格式(GIF),标记图像文件格式(TIFF)和可移植文件格式(PDF)等。

LZW编码在概念上非常简单(Welch[1984])。在编码处理的开始阶段,先构造一个对信源符号进行编码的编码本或"字典"。对8位的单色图像,字典中前256个字被分配给灰度值0,1,2,…,255。当编码器顺序地分析图像像素的时候,字典中没有包括的灰度级序列由算法决定(例如,下边没有用到)其出现的位置。例如,如果图像前两个像素为白色,序列"255-255"可能被分配在位置为256的地方,这个位置下面的地址保留给灰度级0到255。下一次再遇到两个连续的白色像素就用码字256(包含序列255-255的位置的地址)表示它们。如果一个9位512字的字典用于编码处理,初始时用于表示两个像素的(8+8)位码字被一个单一的9位码字所代替。很明显,字典的大小是一个很重要的系统参量。如果字典太小,灰度级序列匹配会变得不太可能;如果太大,码字的尺寸反而会影响压缩性能。

例8.12 一个LZW编码例子

考虑下列4×4、8位图像的垂直边缘:

39	39	126	126
39	39	126	126
39	39	126	126
39	39	126	126

表8.7中详细说明了对它的16个像素进行编码的步骤。假设一个512字带有下列初始内容的字典:

字典位置	条目
0	0
1	1
⋮	⋮
255	255
256	—
⋮	⋮
511	—

位置256到511在开始时是用不到的。

表 8.7　LZW 编码示例

当前识别序列	被处理的像素	编码输出	字典位置(码字)	字典条目
	39			
39	39	39	256	39-39
39	126	39	257	39-126
126	126	126	258	126-126
126	39	126	259	126-39
39	39			
39-39	126	256	260	39-39-126
126	126			
126-126	39	258	261	126-126-39
39	39			
39-39	126			
39-39-126	126	260	262	39-39-126-126
126	39			
126-39	39	259	263	126-39-39
39	126			
39-126	126	257	264	39-126-126
126		126		

图像通过从左到右,从上到下的顺序处理其像素进行编码。每个连续的灰度级值都与一个可变的——表 8.7 的第一列——“当前识别序列”相联系。如所见到的,这个可变序列初始时为空。在字典中搜索每个连续的序列,如果找到了,同表中第一行的情况一样,用新连接和识别(也就是,在字典中定位)的序列代替。表中第二行第一列是进行这样的操作的结果。没有生成输出代码,也没有改变字典。然而,如果没有找到连续序列,则将当前识别序列的地址作为下一个编码值输出,将连续的但未被识别的序列加到字典中,并且当前识别序列被初始化为当前像素的值。表中第二行就是这种情况。最后两列详细说明了在扫描整个 4×4 图像时加到字典中的灰度级序列。在表中定义了 9 个附加的码字。在编码的结尾部分,字典包含了 265 个码字且 LZW 算法已经成功识别了几个重复的灰度级序列——使这些序列从初始时的 128 比特图像减少到 90 比特(也就是,10 个 9 比特编码)。通过从顶部到底部的顺序读取第 3 列的编码得到编码输出。压缩率为 1.42:1。

刚才示范的 LZW 编码的独特特征是编码的字典或码书在对数据进行编码的时候创建的。很明显,一个 LZW 解码器对编码数据流进行解码的同时生成了一个统一的解压缩字典。留下这个例子给读者(见习题 8.16)作为对前面例子的输出进行解码和重构码书的练习。尽管在这个例子中不需要,但在大多数实际应用中都需要一种处理字典溢出的策略。一种简单的解决办法就是,当字典已满并用一个新的初始化的字典进行编码的时候对字典进行刷新或重新初始化。作为一种更为复杂的选择是,监测压缩的性能并在字典变得效能低下或不能接受的时候对字典进行刷新。另一方面,可以对字典中用得最少的条目进行跟踪以便在必要的时候

对其进行替换。

8.4.3 位平面编码

另一种有效的减少像素间冗余的技术就是单独处理图像的位平面。这种技术称为位平面编码。它是以将一幅多级(单色的或彩色的)图像分解为一系列二值图像并通过几种熟知的二值图像压缩方法对每幅二值图像进行压缩的原理为基础的。在这一小节中,将描述使用最为普遍的分解方法,并回顾几种更为普通的用于压缩的方法。

位平面分解

一幅 m 比特的灰度图像具有的灰度级可以用以 2 为底的多项式进行表示:

$$a_{m-1}2^{m-1} + a_{m-2}2^{m-2} + \cdots + a_1 2^1 + a_0 2^0 \tag{8.4.2}$$

基于这种性质,将这类图像分解成一个二值图像集的一种简单方法就是,将多项式的 m 个系数分离到 m 个 1 比特的位平面之中。如在第 3 章中所提到的那样,零级位平面是通过收集每个像素的 a_0 位生成的,面第 $(m-1)$ 级位平面包含 a_{m-1} 比特或系数。一般来讲,每个位平面都根据 0 到 $m-1$ 进行编号,并通过令平面的像素等于每个像素在原图像中对应位的值或多项式的系数进行构造。这种方法固有的缺点是图像在灰度级上稍有变化就会对位平面的复杂性产生显著的影响。例如,如果一个亮度为 127(01111111)的像素与一个亮度为 128(10000000)的像素相邻,每个位平面将包含一个对应 0 到 1(或 1 到 0)的转换。比如,当 127 和 128 的两个二进制编码最高有效位不同的时候,位平面 7 将包含与一个 1 值像素相邻的一个 0 值像素,在这一点产生一次从 0 到 1(或 1 到 0)的转换。

一种可作为替代的分解方法(可以减少小的灰度级变化带来的影响)是首先用一个 m 比特的灰度编码表示图像。这个 m 比特的灰度编码 $g_{m-1}\cdots g_2 g_1 g_0$ 对应式(8.4.2)中的多项式,可以用下列方法计算得到:

$$g_i = a_i \oplus a_{i+1} \qquad 0 \leqslant i \leqslant m-2 \tag{8.4.3}$$

$$g_{m-1} = a_{m-1}$$

这里,$\oplus$表示异或运算。这种编码的惟一特性就是连续码字只在一比特位置上不相同。因此,灰度级小的变化不太可能影响到所有的 m 个位平面。例如,当灰度级 127 和 128 相邻的时候,只有第 7 个位平面含有一次 0 到 1 的转换,因为对应 127 和 128 的灰度编码分别为 11000000 和 01000000。

例 8.13 位平面编码

图 8.14(a)和(b)中所示的 1024×1024 大小的图像用于说明本节其余部分描述的压缩技术。一个小孩的 8 比特单色图像是使用高分辨率 CCD 照相机拍摄的。一份由总统安德鲁·杰克逊于 1796 年写的担保契约的二值图像是在一个文件扫描平台上生成的。图 8.15 和图 8.16 显示了小孩图像中的 8 个二值和灰度编码的位平面。

注意,高阶位平面比它们对应的低阶部分要简单得多。即,这些位平面中包含大块的均匀区域,其中的图像细节相当少、忙乱或随意性较小。另外,灰度编码位平面比对应的二值位平面的复杂性小。

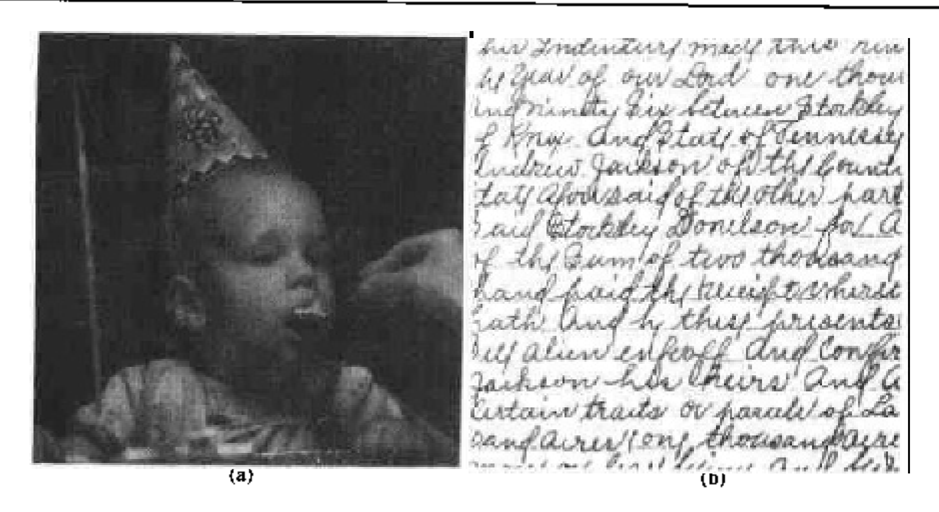


图8.14 一幅1024×1024大小的(a)8比特单色图像和(b)二值图像

常数值区域编码

压缩二值图像或位平面的一种简单但有效的方法是使用指定的码字识别大片连续的1和0区域。这样一种方法称为常数区域编码(CAC),图像被分成大小为$p \times q$个像素的不同块,这些块被分成全白色、全黑色或混合亮度等不同的类。出现可能性最大或最频繁的类被分配给1比特码字0,另两个类分给2比特码10和11。因为通常用于表示每个常数区域的pq比特被1比特或2比特码字代替,所以可以实现压缩。当然,分配给混合亮度类的编码被作为一个前缀使用,这个前缀的后面跟着块的pq比特模式。

当对主要为白色的文本文件进行压缩的时候,稍微简单一些的方法是将纯白色区域的编码定为0,面所有其他的块(包括纯黑色的块)用块的比特模式跟着的编码1进行编码。这种方法充分利用了对被压缩图像的预期结构化倾向而被称为白色色块跳跃(WBS)。由于不希望有太多纯黑色区域存在,因此这些区域都是与混合亮度区域聚合在一起的,允许对高概率的白色色块使用1比特码字。对这一过程(块大小为$1 \times q$)的一种非常有效的改进是用0对纯白色线进行编码,面其他的线使用1后面跟通常的WBS编码序列进行编码。另一种方法是使用迭代的方法将二值图像或位平面不断分解为越来越小的子块。对二维块,一幅纯白色的图像用编码0表示,面其他的图像被分为子块并使用1为前缀进行相似的编码。即,如果一个子块是全白色的,色块的表示将使用前缀1表示此块为第一次迭代的子块,后面跟着编码0,表示色块为纯白色。如果子块不是纯白色的,则分解过程将重复下去直到达到预先定义的子块尺寸,且通过0(如果色块是全白色)或1后面跟着块比特模式进行编码。

一维行程编码

对常数区域进行编码的一种有效的替代方法是用一长度序列表示图像或位平面的每一行,这些长度描绘了对黑色和白色像素的连续行程。这种技术称为行程编码,它是在20世纪50年代发展起来的一种技术,连同这种技术在二维方面的扩展,现已成为传真(FAX)编码的标准压缩方法。基本概念是,对从左到右扫描一行时所遇到的1或0的连接组,使用这些连接组的长度进行编码,并且建立决定行程值的约定。决定行程长度值最通常的方法是:(1)指定每一行第一次行程的值,或(2)假设每一行从白色行程开始,这次行程的长度可能实际上为0。

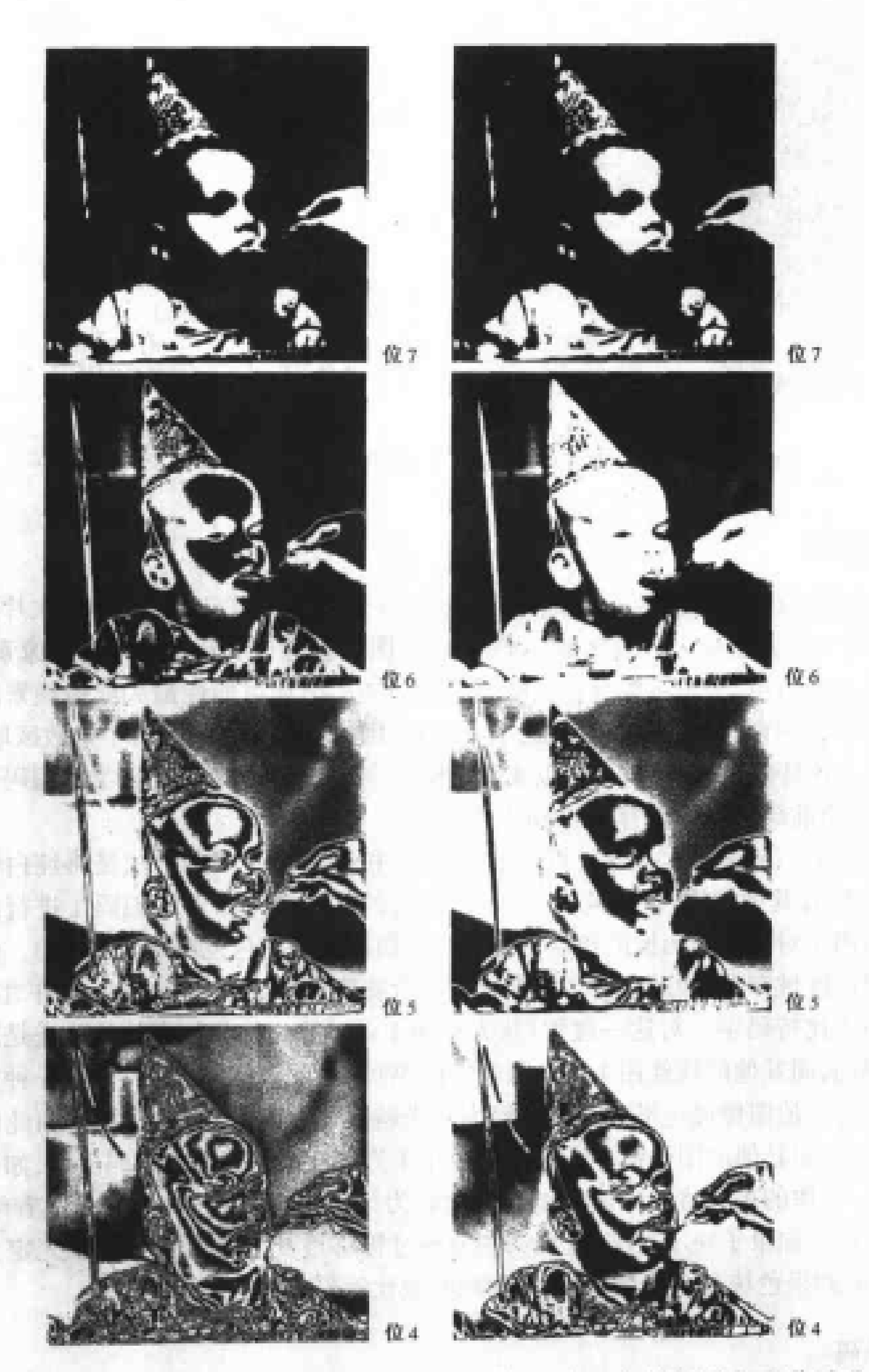


图 8.15　图 8.14(a)中图像的 4 幅最明显的二值(左列)和灰度编码(右列)位平面

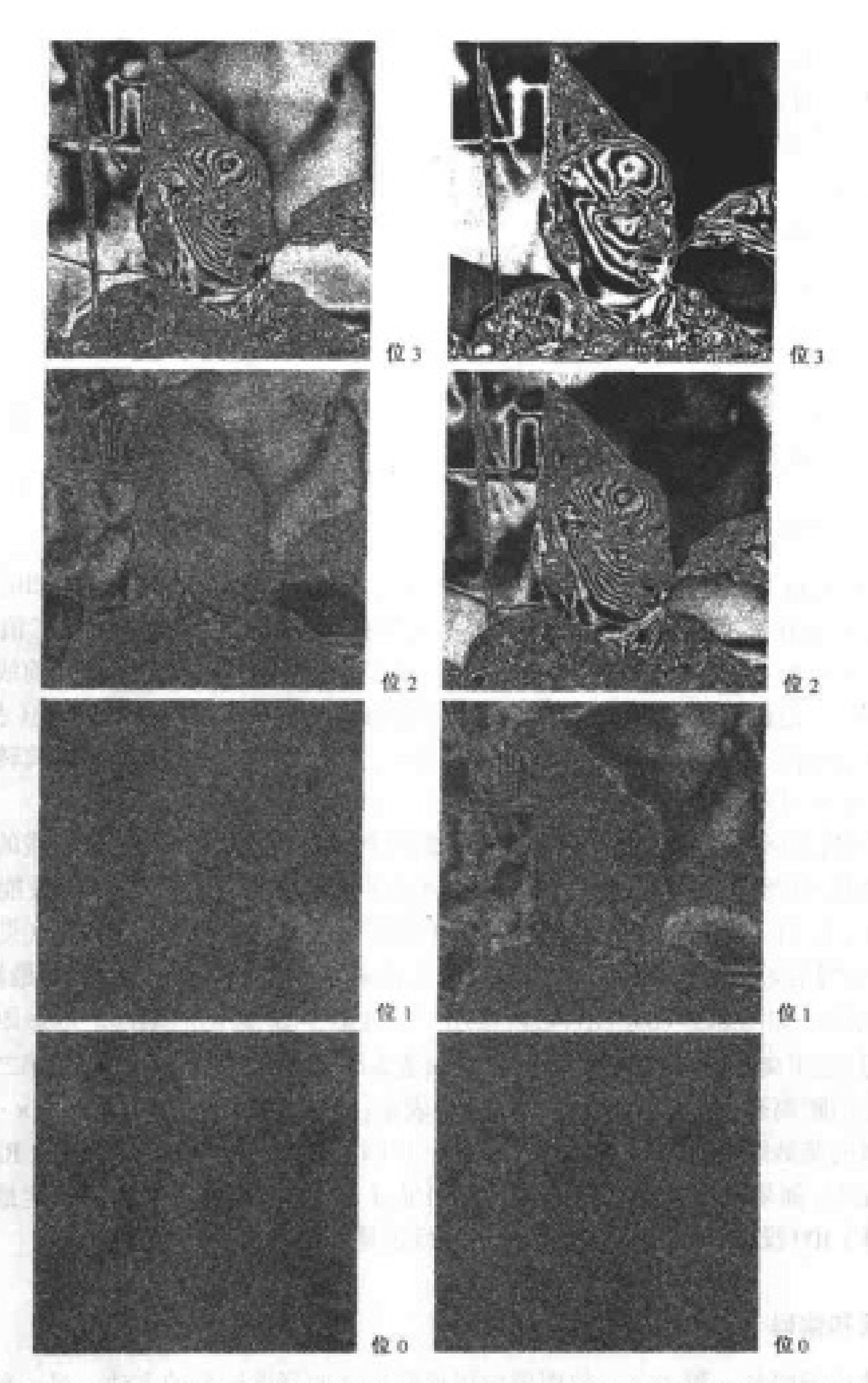


图 8.16　图 8.14(a)中图像的 4 幅最不明显的二值(左列)和灰度编码(右列)位平面

尽管行程长度编码本身是一种压缩图像的有效方法(见 8.1.2 节中的例子),但其他的压缩方法通常可以通过对行程长度本身进行变长编码实现。实际上,黑色和白色的行程长度可以使用变长编码分别进行编码,这些变长编码是根据它们自身的统计数据特殊制定的。例如,符号 a_j 表示一个长度为 j 的黑色行程,可以估计出一个假想的黑色行程长度信源产生的符号 a_j 的概率,而符号 a_j 产生的概率是用整幅图像中长度为 j 的黑色行程的次数除黑色行程的总次数得到的。这种黑色行程长度信源熵的估计表示为 H_0,紧接着把这些概率代入式(8.3.3)。对于白色行程的熵有相似的结果,它表示为 H_1。图像的近似行程长度的熵为:

$$H_{RL} = \frac{H_0 + H_1}{L_0 + L_1} \tag{8.4.4}$$

这里,变量 L_0 和 L_1 分别表示黑色和白色行程长度的平均值。式(8.4.4)提供了对二值图像的行程长度使用变长编码时所要求的每像素平均比特数的估计。

二维行程长度编码

一维行程长度编码概念很容易扩展得到各种二维编码过程。其中较为熟知的一种方法是相对地址编码(RAC)。这种方法是以记录每次黑色和白色行程开始和结束的二值转换的原理为基础的。图 8.17(a)说明了这种方法的实现。注意,ec 是从当前转换 c 到当前线 e 的最后转换的距离,而 cc' 是在前面提到的线上从 c 到首次经过 e 的相似(距离相同)转换(表示为 c')的距离。如果 $ec \leqslant cc'$,则 RAC 编码的距离 d 等于 ec,并且用它表示在 c 处的当前转换。但是如果 $cc' < ec$,则令 d 等于 cc'。

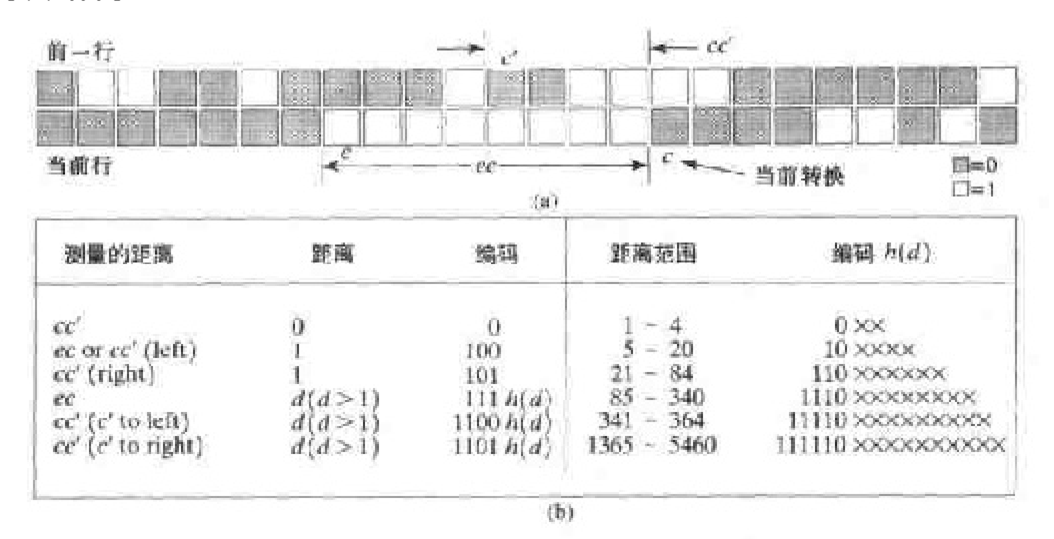
类似行程长度编码,相对地址编码要求按惯例判定行程值。另外,在每条线的起点和终点处假想的转换,以及一条假想的开始线(如,一条全白色的线)必须要考虑,以便能够适当处理图像的边界。最后,由于多数图像的 RAC 距离的概率分布实际上是不均匀的(见 8.1.1 节),因此,RAC 处理的最后步骤是通过使用适当的变长编码对选定的 RAC 距离(即最短距离)和距离 d 进行编码。如图 8.17(b)所示,可以使用一种近似于 B_1 编码的编码。最短距离分配给最短码字,而其他距离是这样编码的:使用一种前缀表示最短的 RAC 距离,使用第二前缀指定 d 在一个特定的距离范围内取值,用 d 的二进制表示[图 8.17(b)中表示为 $\times\times\times\cdots\times$]减去取值范围本身的基数距离。如果 ec 和 cc' 如图 8.17(a)所示为 +8 和 +4,正确的 RAC 码字应为 1100011。最后,如果 $d=0$,c 直接在 c' 下,而如果 $d=1$,则解码器可能必须判定最接近的转换点,因为编码 100 没有指明测量是相对于当前行还是相对于前一行进行的。

轮廓线追踪和编码

相对地址编码是一种表示二值图像中组成轮廓线的亮度转换的方法。另一种方法是通过一系列的边界点或通过一个单一的边界点和一个方向集合表示每条轮廓线。后一种方法有时称为直接轮廓跟踪法。在这一节中,将同时讲解另一种称为预测微分量化(PDQ)的方法。这种方法证明了上述两种方法的本质特点。它是一种面向扫描线的轮廓跟踪过程。

在预测微分量化中,对图像每个对象的前部和后部的轮廓线(图 8.18)同时进行跟踪以生成一个序列对(Δ',Δ'')。Δ'项是邻近线上前部轮廓的起始坐标之间的差。Δ''是从前到后的轮廓线长度的差。这些差同表示新轮廓线的开始(新的开始信息)和旧轮廓线的终结(聚合信息)

的特定信息一起表示每个对象。如果 Δ″用相邻线的后轮廓坐标之间的差(表示为 Δ‴)替代，则这种技术称为双增量编码(DDC)。



测量的距离	距离	编码	距离范围	编码 $h(d)$
cc'	0	0	1 - 4	0 ××
ec or cc' (left)	1	100	5 - 20	10 ××××
cc' (right)	1	101	21 - 84	110 ××××××
ec	$d(d>1)$	111 $h(d)$	85 - 340	1110 ××××××××
cc' (c' to left)	$d(d>1)$	1100 $h(d)$	341 - 364	11110 ××××××××××
cc' (c' to right)	$d(d>1)$	1101 $h(d)$	1365 - 5460	111110 ××××××××××××

(b)

图 8.17 一种相对地址编码(RAC)的说明

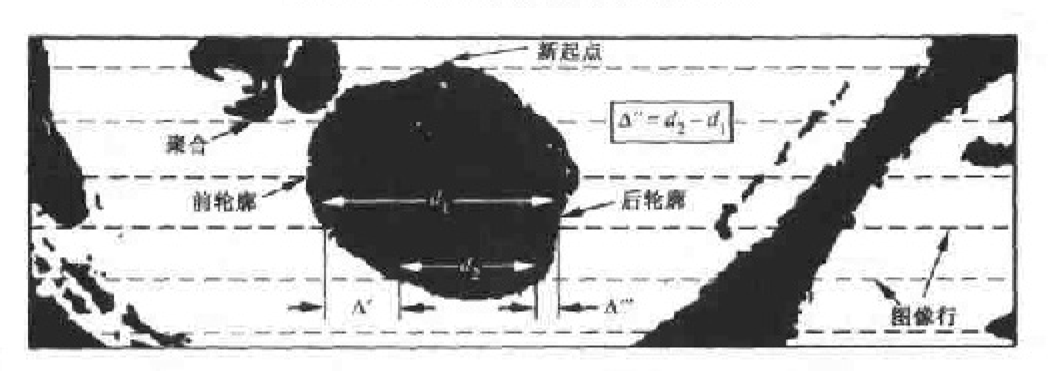


图 8.18 PDQ 算法的参数

新的开始和聚合信息允许在扫描线基础上产生(Δ′,Δ″)或(Δ′,Δ‴)，适当地连接到前一行和下一行中相对应的偶对。没有这些信息，解码器无法对另一个偶对参考一个差分偶对，或在图像中正确地放置轮廓线。为了避免对每个新的起点和聚合信息的行列坐标同时进行编码，通常使用惟一的编码识别不包含物体像素的扫描线。PDQ 和 DDC 编码过程的最后一步是用适当的变长编码表示 Δ′,Δ″或 Δ‴和新的起点与聚合的坐标。

例 8.14 二值压缩技术的比较

通过对前述二值压缩技术进行对比来对这一节进行总结。每种方法都用于压缩图 8.14 中的图像。得到的编码率和压缩率在表 8.8 和表 8.9 中列出。解释这些结果的时候，注意关于 RLC 行程长度的熵的一阶估计(见 8.3.4 节)及 PDQ 和 DDC 距离的计算和使用，是作为在用 8.4.1 节中的变长编码方法时所能达到的压缩性能的一种近似。

表 8.8 和表 8.9 中的结果说明所有的技术均能消除一部分像素间冗余。即，得到的编码率小于每幅图像的一阶熵的估计。行程编码被证明是最好的位平面编码图像的编码方法，而二维技术(比如 PDQ，DDC 和 RAC)在压缩二值图像时表现出更好的性能。再有，图 8.14(a)中图像的灰度编码相对简单的过程改进的编码性能在 1 比特/像素左右。最

后,注意这5种压缩方法都可以使用一个1到2之间的因子对单色图像进行压缩,而对图8.14(b)的二值图像的压缩使用2到5之间的一个因子。如表8.8所示,产生这种性能差异的原因是算法无法对位平面编码图像中的较低阶的位平面进行压缩。实际上,表中画横线的项表明算法引起了数据量膨胀。在这些情况下,用原始数据表示位平面,总共的编码率只增加了1比特/像素。

表8.8 图8.14(a)无误差位平面编码结果:$H \approx 6.82$ 比特/像素

方法	位平面编码率(比特/像素)								编码率	压缩率
	7	6	5	4	3	2	1	0		
二值位平面编码										
CBC(4×4)	0.14	0.24	0.60	0.79	0.99	—	—	—	5.75	1.4:1
RLC	0.09	0.19	0.51	0.68	0.87	1.00	1.00	1.00	5.33	1.5:1
PDQ	0.07	0.18	0.79	—	—	—	—	—	6.04	1.3:1
DDC	0.07	0.18	0.79	—	—	—	—	—	6.03	1.3:1
RAC	0.06	0.15	0.62	0.91	—	—	—	—	5.17	1.4:1
灰度位平面编码										
CBC(4×4)	0.14	0.18	0.48	0.40	0.61	0.98	—	—	4.80	1.7:1
RLC	0.09	0.13	0.40	0.33	0.51	0.85	1.00	1.00	4.29	1.9:1
PDQ	0.07	0.12	0.61	0.40	0.82	—	—	—	5.02	1.6:1
DDC	0.07	0.11	0.61	0.40	0.81	—	—	—	5.00	1.6:1
RAC	0.06	0.10	0.49	0.31	0.62	—	—	—	4.05	1.8:1

表8.9 图8.14(b)的无误差二值图像压缩结果:$H \approx 0.55$ 比特/像素

	WBS(1×8)	WBS(4×4)	RLC	PDQ	DDC	RAC
编码率(比特/像素)	0.48	0.39	0.32	0.23	0.22	0.23
压缩率	2.1:1	2.6:1	3.1:1	4.4:1	4.7:1	4.4:1

8.4.4 无损预测编码

现在转向无误差压缩方法,这种方法不需要将图像分解为一个位平面的集合。这种方法通常称为无损预测编码,它是基于通过对每个像素新增的信息进行提取和编码,来消除在空间上较为接近像素之间的冗余信息。一个像素的新增信息被定义为此像素实际值和预测值之间的差异。

图8.19显示了一个无损预测编码系统的基本组成部分。这个系统由一个编码器和一个解码器组成,每部分都包含一个相同的预测器。由于输入图像的每个表示为 f_n 的连续像素都要送入编码器,预测器根据以往的一些输入生成输入像素的预期值。预测器的输出被四舍五入为最接近的整数 $\hat{f}_n$,并将这个整数用于计算差异或预测误差:

$$e_n = f_n - \hat{f}_n \tag{8.4.5}$$

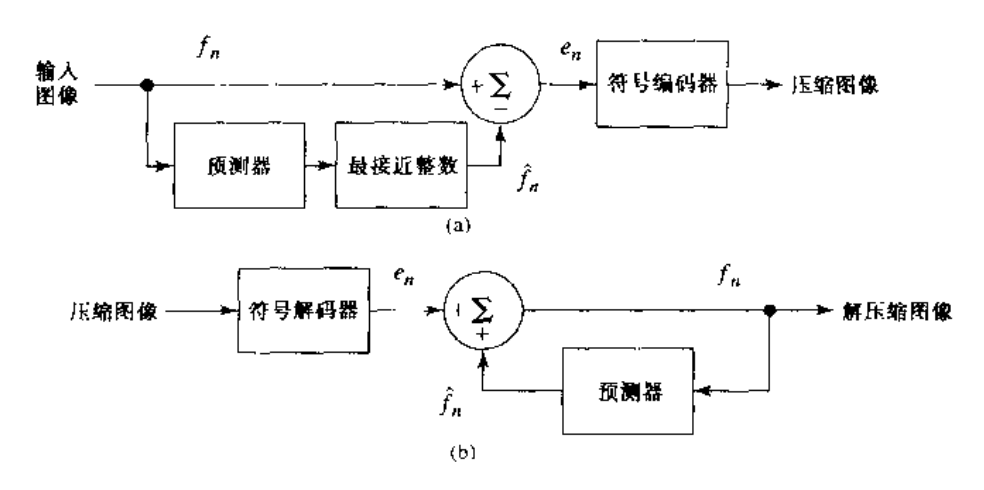


图 8.19 一个无损预测编码模型。(a)编码器,(b)解码器

这个误差使用变长代码(通过符号编码器)进行编码以生成压缩数据流的下一个元素。图 8.19(b)中的解码器根据接收到的变长码字对 e_n 进行重构并实现反运算:

$$f_n = e_n + \hat{f}_n \tag{8.4.6}$$

有多种局部的、全局的和自适应(见8.5.1节)的方法可以用于生成 $\hat{f}_n$。然而,在大多数情况下,预测是根据前 m 个像素的线性组合生成的。即,

$$\hat{f}_n = \text{round}\left[\sum_{i=1}^{m} \alpha_i f_{n-i}\right] \tag{8.4.7}$$

这里,m 是线性预测器的阶,round 是表示四舍五入或取最接近的整数的运算的函数,α_i($i=1,2,\cdots,m$)是预测系数。在光栅扫描应用中,下标 n 根据它们的发生时间指示预测器输出。即,式(8.4.5)到式(8.4.7)中的和可以用更明确的符号和替代,这里 t 表示时间。在其他情况下,n 用做图像的空间坐标和/或帧数(某一时间内的图像序列)的下标。例如,在一维线性预测编码中,式(8.4.7)可以写为:

$$\hat{f}_n(x,y) = \text{round}\left[\sum_{i=1}^{m} \alpha_i f(x, y-i)\right] \tag{8.4.8}$$

这里每个下标变量都被直接表示为一个空间坐标 x 和 y 的函数。注意,式(8.4.8)表明一维线性预测 $\hat{f}(x,y)$ 仅是一个当前行上前几个像素的函数。在二维预测编码中,预测是一个在从左到右、从上到下扫描图像过程中前几个像素的函数。在三维情况下,预测是以这些像素和前面几帧的像素为基础的。式(8.4.8)不能对每一行的前 m 个像素进行求解,因此这些像素必须用其他的方法进行编码(比如,霍夫曼编码)并被视为预测编码处理的额外开销。相似的方法可用于更高维的情况。

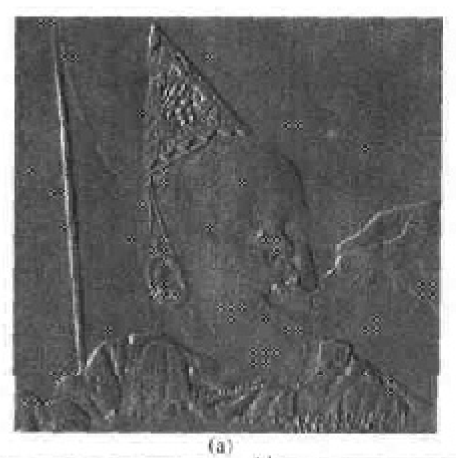
例 8.15 预测编码

考虑使用简单的一阶线性预测器对图 8.14(a)的单色图像进行编码:

$$\hat{f}_n(x,y) = \text{round}[\alpha f(x, y-1)] \tag{8.4.9}$$

这种一般形式的预测器称为前像素预测器,相应的预测编码过程称为差分编码或前像素编码。图 8.20(a)显示了根据式(8.4.9)令 $\alpha=1$ 时得到的预测误差图像。在这幅图像中,灰度级 128 表示一个零的预测误差,而所有非零的正的和负的预测误差(在估算范围内的

和超出估算范围的)都与 8 相乘并分别被显示成更亮的和更暗的灰度区。预测图像的均值为 128.02,这个值对应的平均预测误差仅为 0.02 比特。



(a)

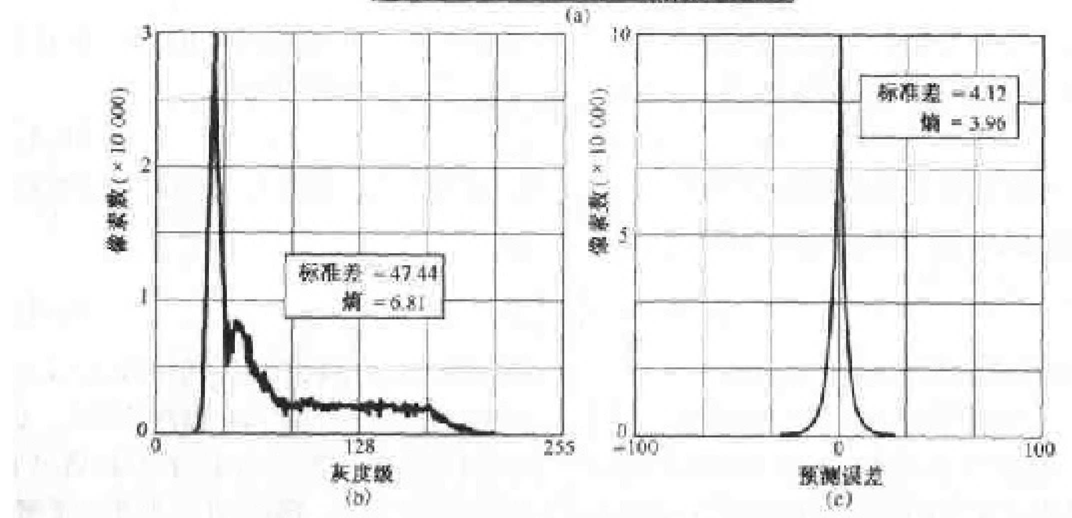


图 8.20 (a)根据式(8.4.9)得到的预测误差图像,(b)原图像的灰度级直方图,(c)预测误差的直方图

图 8.20(b)和(c)显示了图 8.14(a)中图像的灰度直方图和根据式(8.4.9)得到的预测误差图像的直方图。注意,图 8.20(c)中的预测误差的变化比原图像中灰度级的变化小得多。另外,预测误差的熵的一阶估计明显小于对应的原图像的一阶估计(3.96 比特/像素对 6.81比特/像素)。这个熵的减少反映了通过预测编码处理消除了大量的冗余,尽管事实上对 m 比特的图像需要$(m+1)$比特数据准确表示根据式(8.4.5)得到的误差序列。虽然 8.4.1 节的任何变长编码过程都可以用于对这个误差序列进行编码,但得到的压缩结果被限制在大约 8/3.96 或约 2:1 的水平。通常,任何无损预测编码方法的最大压缩估计,可以通过将用于表示原始图像中每一个像素的平均比特数除以预测误差数据的熵的一阶估计而得到。

前面的例子重点强调了,用无损预测编码实现的压缩量与将输入图像映射到预测误差序列得到的熵的减少直接相关。

因为通过预测和差分处理,消除了大量的像素间冗余,所以,预测误差的概率密度函数通常在零处有一个很高的峰,并表现出变化相对较小(同输入灰度级分布相比较而言)的特征。

实际上，预测误差的密度函数通常使用零均值不相关的拉普拉斯概率密度函数进行模型化：

$$p_e(e) = \frac{1}{\sqrt{2}\sigma_e} e^{\frac{-\sqrt{2}|e|}{\sigma_e}} \tag{8.4.10}$$

这里 σ_e 是 e 的标准差。

8.5 有损压缩

与前一节讲述的无误差方法不同，有损编码是以在图像重构的准确度上做出让步而换取压缩能力增加的概念为基础的。如果产生的失真(可能是明显可见的也可能不很明显)是可以容忍的，则压缩能力上的增加就是有效的。实际上，许多种有损编码技术有能力根据压缩比率超过 100:1 的数据重构实际上不可区分的单色图像，并且生成的图像与对原图进行 10:1 到 50:1压缩的图像之间没有本质上的区别。单色图像的无差错编码很少能在数据压缩上得到大于3:1的结果。正如 8.2 节中指明的，这两种方法之间主要的差别在于，是否存在图 8.6 中的量化器模块。

8.5.1 有损预测编码

这一节，在 8.4.4 节介绍的模型中增加一个数字量化器，并分析在重构准确度和压缩性能之间得到的折中。如图 8.21 所显示的，量化器充当无误差编码器的取整函数，它被插入在符号编码器和形成预测误差的那一点之间。

这个量化器将预测误差映射成有限范围内的输出，表示为 $\dot{e}_n$，这个输出确定了与有损预测编码相联系的压缩和失真的量。

为了适应量化步骤的加入，必须更改图 8.19(a)中的无误差编码器，以便由编码器和解码器生成的预测相等。如图 8.21(a)所示，这是通过在反馈环中设置一个有损编码器的预测器来实现的，这里，表示为 $\dot{f}_n$ 的反馈环的输入是过去预测函数和对应的量化误差产生的。即，

$$\dot{f}_n = \dot{e}_n + \hat{f}_n \tag{8.5.1}$$

这里，$\hat{f}_n$ 同 8.4.4 节中所定义的一样。这个闭合的循环结构可以防止在解码器的输出部位形成误差。注意，从图 8.21(b)可以看出，解码器的输出也是由式(8.5.1)给出的。

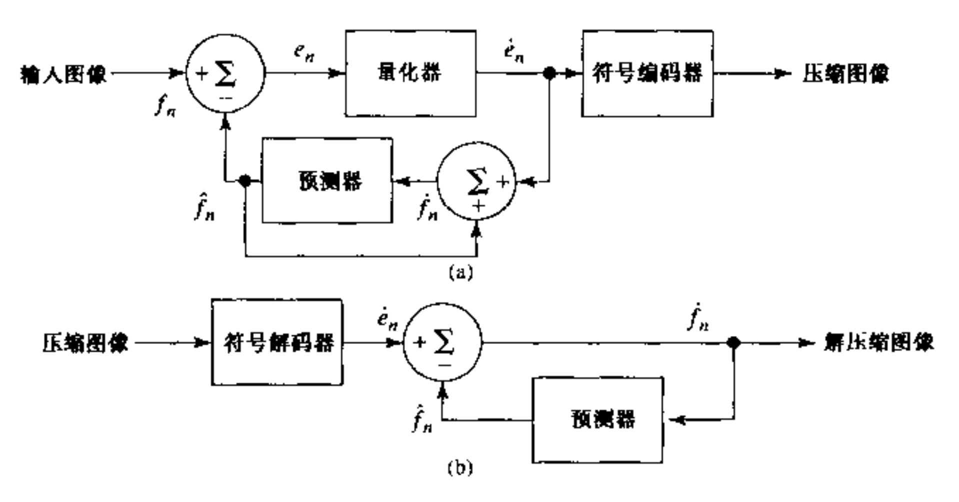


图 8.21　有损预测编码模型。(a)编码器和(b)解码器

例 8.16 德尔塔调制

德尔塔调制(DM)是一种简单但众所周知的有损预测编码形式,其中,预测器和量化器定义为:

$$\hat{f}_n = \alpha \dot{f}_{n-1} \tag{8.5.2}$$

和

$$\dot{e}_n = \begin{cases} +\zeta & e_n > 0 \\ -\zeta & \text{其他} \end{cases} \tag{8.5.3}$$

这里,α 是一个预测系数(通常小于 1),而 ζ 是一个正的常量。量化器的输出 $\dot{e}_n$ 可以用 1 比特表示[图 8.22(a)],因此,图 8.21(a)的符号编码器可以使用 1 比特的固定长度编码。得到的 DM 码率是 1 比特/像素。

图 8.22(c)说明了调制处理的机理,这里的计算要求压缩和重构被列入表格的输入序列{14,15,14,15,13,15,15,14,20,26,27,28,27,27,29,37,47,62,75,77,78,79,80,81,81,82,82},$\alpha = 1$,$\zeta = 6.5$。这个处理过程从第一个输入像素无误差地传递到解码器开始。由于在编码器和解码器都设置了初始条件 $\dot{f}_0 = f_0 = 14$,余下的输出可以通过反复计算式(8.5.2)、式(8.4.5)、式(8.5.3)和式(8.5.1)得到。这样,例如,当 $n = 1$ 时,$\hat{f} = (1)(14) = 14$,$e_1 = 15 - 14 = 1$,$\dot{e}_1 = +6.5$(因为 $e_1 > 0$),$\dot{f}_1 = 6.5 + 14 = 20.5$,得到的重构误差是(15 - 20.5)或 -5.5灰度级。

图 8.22(b)显示了图 8.22(c)中显示的图表化的表格数据。其中显示了输入和完整的解码输出(f_n 和$\dot{f}_n$)。注意,在从 $n = 14$ 到 19 这一段快速变化的区域中,ζ 太小不足以表示输入的最大变化,从而产生了被称为斜率过载的失真。另外,当 ζ 太大无法表示输入的最小变化的时候,如在从 $n = 0$ 到 $n = 7$ 的相对平滑的区域中,会出现颗粒噪声。在大多数图像中,这两种现象会导致图像的边缘变得模糊,呈现粒状或有噪声的表面(即,失真的平滑区域)。

在前例中提到的失真对所有有损预测编码方法都是很普遍的。这些失真的严重性取决于所使用的量化和预测方法之间的一个相互作用的复杂集合。不考虑它们相互间的作用,预测器通常是在无量化误差的假设之下设计的,且量化器被设计成其自身的误差最小。就是说,预测器和量化器是彼此独立进行设计的。

最佳预测器

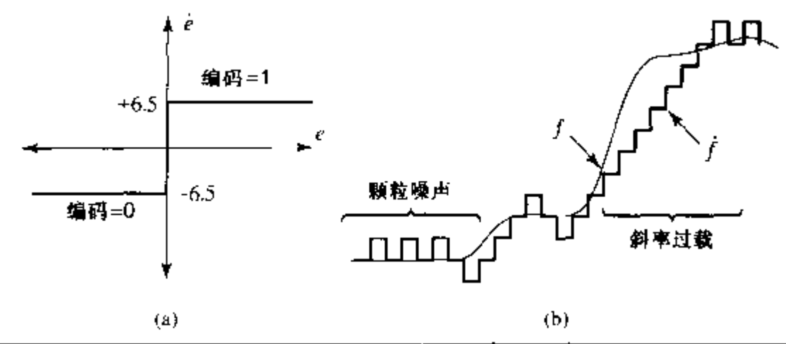
在大多数预测编码应用中的最佳预测器可将编码器的均方预测误差降至最小[①]:

$$E\{e_n^2\} = E\{[f_n - \hat{f}_n]^2\} \tag{8.5.4}$$

限制条件为:

$$\dot{f}_n = \dot{e}_n + \hat{f}_n \approx e_n + \hat{f}_n = f_n \tag{8.5.5}$$

① 符号 $E\{\cdot\}$代表期望算子。



输入		编码器				解码器		误差
n	f	$\hat{f}$	e	$\dot{e}$	$\dot{f}$	$\hat{f}$	$\dot{f}$	$[f-\dot{f}]$
0	14	—	—	·	14.0	—	14.0	0.0
1	15	14.0	1.0	6.5	20.5	14.0	20.5	-5.5
2	14	20.5	-6.5	-6.5	14.0	20.5	14.0	0.0
3	15	14 0	1.0	6.5	20.5	14.0	20.5	-5.5
·	·	·	·	·	·	·	·	·
·	·	·	·	·	·	·	·	·
14	29	20.5	8.5	6.5	27.0	20.5	27.0	2.0
15	37	27.0	10.0	6.5	33.5	27.0	33.5	3.5
16	47	33.5	13.5	6 5	40.0	33.5	40.0	7.0
17	62	40.0	22.0	6.5	46.5	40.0	46.5	15.5
18	75	46.5	28.5	6.5	53.0	46.5	53.0	22.0
19	77	53.0	24 0	6.5	59.6	53.0	59.6	17.5
·	·	·	·	·	·	·	·	·
·	·	·	·	·	·	·	·	·

(c)

图 8.22 德尔塔调制的一个例子

和

$$\hat{f}_n = \sum_{i=1}^{m} \alpha_i f_{n-i} \tag{8.5.6}$$

即,选择最小化均方预测误差为最佳的准则。量化误差假设是可以忽略的($\dot{e}_n \approx e_n$),且预测被限定为前 m 个像素的线性组合[①]。这些限制并不是基本的,但它们在相当大程度上简化了分析,并且同时减小了预测器计算的复杂性。如此得到的预测编码方法被称为差分脉冲编码调制(DPCM)。

在这些条件下,最佳预测器的设计问题变成相对简单的选择 m 个预测系数的应用,这些系数使下式最小:

$$E\{e_n^2\} = E\left\{\left[f_n - \sum_{i=1}^{m} \alpha_i f_{n-i}\right]^2\right\} \tag{8.5.7}$$

相对于每个系数对式(8.5.7)求导,令导数为零,在假设 f_n 具有零均值和方差为 σ^2 的条件下解出联立方程的解集,得到:

$$\boldsymbol{\alpha} = \mathbf{R}^{-1}\mathbf{r} \tag{8.5.8}$$

这里 $\mathbf{R}^{-1}$ 是下列 $m \times m$ 自相关矩阵的逆矩阵:

① 一般来讲,一个非高斯场的图像的最佳预测器是用于估计像素的一个非线性函数。

$$\mathbf{R}=\begin{bmatrix} E\{f_{n-1}f_{n-1}\} & E\{f_{n-1}f_{n-2}\} & \cdots & E\{f_{n-1}f_{n-m}\} \\ E\{f_{n-2}f_{n-1}\} & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & & & \\ E\{f_{n-m}f_{n-1}\} & E\{f_{n-m}f_{n-2}\} & \cdots & E\{f_{n-m}f_{n-m}\} \end{bmatrix} \tag{8.5.9}$$

$\mathbf{r}$ 和 $\boldsymbol{\alpha}$ 是 m 元向量:

$$\mathbf{r}=\begin{bmatrix} E\{f_nf_{n-1}\} \\ E\{f_nf_{n-2}\} \\ \vdots \\ E\{f_nf_{n-m}\} \end{bmatrix}, \quad \boldsymbol{\alpha}=\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix} \tag{8.5.10}$$

因此,对任何输入图像,使式(8.5.7)取得最小的系数可以通过一系列基本的矩阵操作确定。另外,这些系数只取决于原始图像中像素的自相关。使用这些最佳系数得到的预测误差的方差为:

$$\sigma_e^2=\sigma^2-\boldsymbol{\alpha}^T\mathbf{r}=\sigma^2-\sum_{i=1}^{m}E\{f_nf_{n-i}\}\alpha_i \tag{8.5.11}$$

尽管计算式(8.5.8)的机理非常简单,但构造 $\mathbf{R}$ 和 $\mathbf{r}$ 需要的自相关的计算实现起来是非常困难的。因此,在实践中,几乎从不使用局部预测(这些预测的预测系数是逐图像进行计算的)。

在大多数情况下,通过假设一个简单的图像模型并将相应的自相关代入式(8.5.9)和式(8.5.10)去计算一系列全局系数。比如,当假设一个带有可分离的自相关函数的二维马尔可夫信源(见 8.3.3 节):

$$E\{f(x,y)f(x-i,y-i)\}=\sigma^2\rho_v^i\rho_h^j \tag{8.5.12}$$

和广义的四阶线性预测器:

$$\begin{aligned}\hat{f}(x,y)=&\ \alpha_1f(x,y-1)+\alpha_2f(x-1,y-1)\\ &+\alpha_3f(x-1,y)+\alpha_4f(x-1,y+1)\end{aligned} \tag{8.5.13}$$

的时候,得到的最佳系数是:

$$\alpha_1=\rho_h \quad \alpha_2=-\rho_v\rho_h \quad \alpha_3=\rho_v \quad \alpha_4=0 \tag{8.5.14}$$

这里,ρ_h 和 ρ_v 分别是所研究的图像的水平和垂直相关系数。

最后,式(8.5.6)中预测系数的和通常要小于或等于 1。即,

$$\sum_{i=1}^{m}\alpha_i\leqslant 1 \tag{8.5.15}$$

这种限制是为了确保预测器的输出能落到灰度级的允许范围内,并减少传输噪声的影响,传输噪声的影响通常在重构图像中表现为水平的条纹。减小 DPCM 解码器对输入噪声的敏感性是很重要的,因为单个差错(在适当的环境下)会传播到所有以后的输出。即,解码器的输出会变得不稳定。通过进一步限制式(8.5.15)必须严格小于 1,输入差错的影响被限制在少量的输出中。

例 8.17　预测技术的对比

考虑由对图 8.23 中的单色图像进行 DPCM 编码产生的预测误差,假设量化误差为零且用

下列四个预测器的一个：

$$\hat{f}(x,y) = 0.97f(x,y-1) \tag{8.5.16}$$

$$\hat{f}(x,y) = 0.5f(x,y-1) + 0.5f(x-1,y) \tag{8.5.17}$$

$$\hat{f}(x,y) = 0.75f(x,y-1) + 0.75f(x-1,y) - 0.5f(x-1,y-1) \tag{8.5.18}$$

$$\hat{f}(x,y) = \begin{cases} 0.97f(x,y-1) & \Delta h \leqslant \Delta v \\ 0.97f(x-1,y) & 其他 \end{cases} \tag{8.5.19}$$

这里，$\Delta h = |f(x-1,y) - f(x-1,y-1)|$ 和 $\Delta v = |f(x,y-1) - f(x-1,y-1)|$ 表示在点(x,y)的水平和垂直梯度。式(8.5.16)到式(8.5.18)定义了一个 α_i 的相对鲁棒性的集合。这个鲁棒性的集合在图像很宽的范围上提供了令人满意的性能。设计式(8.5.19)的自适应预测器，通过计算图像的方向特性的局部度量(Δh 和 Δv)，并选择适合度量性能的预测器来改善边缘的重现质量。



图 8.23 一幅 512×512 大小的 8 比特单色图像

图 8.24(a)到(d)显示了预测误差图像，这些图像是使用式(8.5.16)到式(8.5.19)的预测器得到的。注意随着预测器阶数的增加明显可见的误差减少了①。预测误差分布的标准差遵从相似的模式。它们分别是 4.9，3.7，3.3 和 4.1 个灰度级。

最佳量化

图 8.25 中所示的阶梯量化函数 $t = q(s)$是关于 s 的奇函数[即，$q(-s) = -q(s)$]，这个函数完全可以用在图形的第一象限所显示的 s_i 和t_i 的$L/2$ 值描述。这些断点定义了函数的不连续性，并被称为量化器的判决和重构级。为了方便起见，如果 s 在半开区间$(s_i, s_{i+1}]$内，则可以将 s 看做到 t_i 的映射。

① 用多于前 3 个或前 4 个像素的预测因子，在预测因子的复杂性增加时无法获得更大的压缩增益(Habibi[1971])。

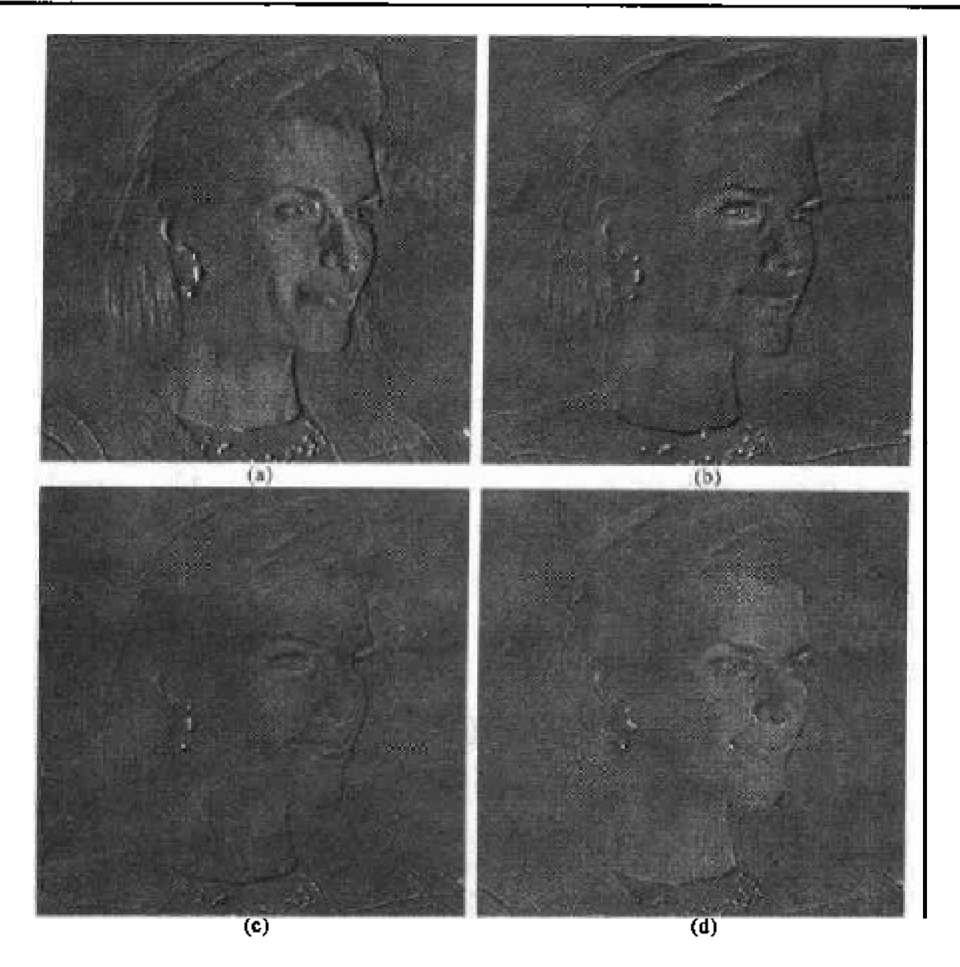


图 8.24 4 种线性预测技术的对比

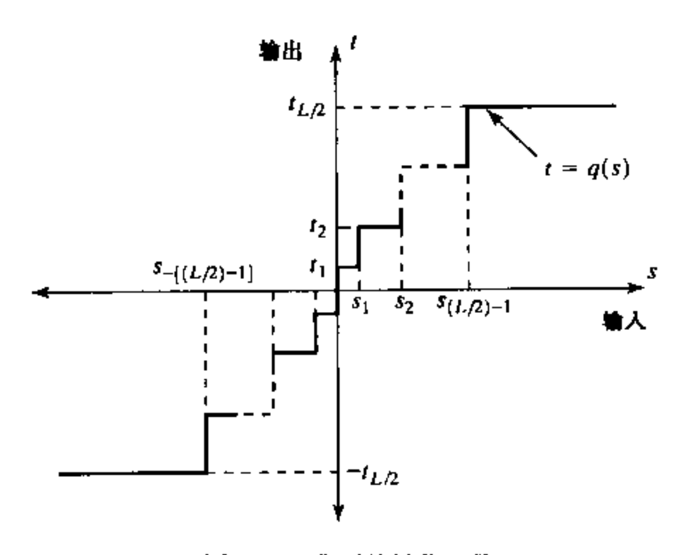


图 8.25 典型的量化函数

量化器的设计问题是，对于特殊的优化准则和输入概率密度函数 $p(s)$ 选择最佳的 s_i 和 t_i。如果最佳化准则(可以是统计上的或心理视觉量度上的①)是均方量化误差最小，并且 $p(s)$

① 参见 Netravali[1977]及 Limb 和 Rubinstein[1978]可了解更多的关于心理视觉量度的内容。

是偶函数,则对于最小误差(Max[1960])的条件是:

$$\int_{s_{i-1}}^{s_i} (s - t_i)p(s)\mathrm{d}s = 0 \quad i = 1,2,\cdots,\frac{L}{2} \tag{8.5.20}$$

$$s_i = \begin{cases} 0 & i = 0 \\ \dfrac{t_i + t_{i+1}}{2} & i = 1,2,\cdots,\dfrac{L}{2} - 1 \\ \infty & i = \dfrac{L}{2} \end{cases} \tag{8.5.21}$$

和

$$s_{-i} = -s_i \quad t_{-i} = -t_i \tag{8.5.22}$$

式(8.5.20)指明重构层是在概率密度函数 $p(s)$ 下超出特定判定区间的区域的质心,而式(8.5.21)指出判定层在重构层之间的一半处。式(8.5.22)是由于 q 为一个奇函数的结果。对于任意的 L,满足式(8.5.20)到式(8.5.22)的 s_i 和 t_i 在均方误差意义下是最佳的;相应的量化器称为 L 层的劳埃德·马克斯(Lloyd-Max)量化器。

表8.10列出了单位方差拉普拉斯概率密度函数[见式(8.4.10)]的2层、4层和8层劳埃德·马克斯判定和重构层。因为对大多数非无效的 $p(s)$ 来说,求取式(8.5.20)到式(8.5.22)的明确的或闭合形式的解是困难的,所以这些值都是数字的(Paez 和 Glisson[1972])。所示的这3个量化器分别提供了1,2和3比特/像素的固定的输出速率。由于表8.10列出的是单位方差分布,所以 $\sigma \neq 1$ 的情况下,重构和判定层用概率密度函数的标准差乘以列表中的值得到。表中的最后一行列出了步长 θ,步长 θ 同时满足式(8.5.20)到式(8.5.22)和附加的限制条件:

$$t_i - t_{i-1} = s_i - s_{i-1} = \theta \tag{8.5.23}$$

如果使用变长码的符号编码器作为通常的图8.21(a)的有损预测编码器,则步长为 θ 的最佳均匀量化器提供的编码率(用于拉普拉斯 pdf),将低于具有相同输出保真度的固定长度编码的劳埃德·马克斯量化器(O'Neil[1971])。

表8.10　单位方差的拉普拉斯概率密度函数的劳埃德·马克斯量化器

层次数	2		4		8	
i	s_i	t_i	s_i	t_i	s_i	t_i
1	∞	0.707	1.102	0.395	0.504	0.222
2			∞	1.810	1.181	0.785
3					2.285	1.576
4					∞	2.994
θ	1.414		1.087		0.731	

尽管劳埃德·马克斯和最佳均匀量化器不是自适应的,但由基于图像的局部状态调整量化层次可以得到更多的好处。在理论上,缓慢变化的区域的量化比较精细,而快速变化的区域的

量化却比较粗糙。这个方法同时可以减少颗粒状噪声和斜率过载,而在编码率上仅要求最小的增加量。代价是增加了量化器的复杂性。

例 8.18　量化和重构的说明

图 8.26(a),(c)和(e)显示了 DPCM 重构图像,这幅图像是由联合使用表 8.10 中的 2 层、4 层和 8 层劳埃德·马克斯量化器用式(8.5.18)的平面预测器得到的。量化器是由列表中的劳埃德·马克斯判定与重构层和从前面例子中得到的非量化的平面预测误差的标准差(即3.3个灰度级)相乘生成的。注意,由于斜率过载,使解码图像的边缘模糊了。在图 8.26(a)中这一点特别显著,它是由两层量化器产生的,但不如在图 8.26(c)和(e)中(它使用了 4 层和 8 层量化器)明显。图 8.27(a),(c)与(e)显示了这些经过解码的图像和图 8.23的原图像之间标定后的差异。

为了生成图 8.26(b),(d)和(f)中的解码图像,并得到图 8.27(b),(d)和(f)中的误差图像,使用自适应量化方法。在这种方法中,对于每 16 个像素的块可以在 4 个可能的量化器中选择最好的(在均方误差意义上)。这 4 个量化器是前述的最佳劳埃德·马克斯量化器的标定过的版本。标定因子为 0.5,1.0,1.75 和 2.5。因为在每个块上附加了两位编码以便确定选择的量化器,所以与量化器开关相联系的开销为 2/16 或 0.125 比特/像素。注意,由此而导致感觉误差上的显著减少,在编码率上则增加得相对较小。

表 8.11 列出了图 8.27(a)到(f)中差异图像的 rms 误差,连同一些预测器和量化器的组合的数字。注意,在均方误差意义上,两层自适应量化器的表现同四层非自适应量化器相近。另外,四层自适应量化器要强于八层非自适应方法。通常,数字上的结果表明式(8.5.15)、式(8.5.17)和式(8.5.19)的预测器表现出来的所有特点与式(8.5.18)的预测器的特点是一样的。使用每种量化方法得到的压缩效果列在表 8.11 的最后一行。注意,由于使用自适应方法产生的 rms 误差[式(8.1.8)]实质性的减少没有对压缩性能产生明显的影响。

8.5.2　变换编码

8.5.1 节中讨论的预测编码技术直接对图像的像素进行操作,因而是空间域的方法。在这一节中,将考虑基于改进图像变换的压缩技术。在变换编码中,一种可逆线性变换(比如,傅里叶变换)用于将图像映射到变换系数集,然后,这些系数被量化和编码。对大多数自然图像,大量系数的量级很小,可以进行不很精确的量化(或完全丢弃),几乎不会产生多少图像失真。包括第 4 章的离散傅里叶变换(DFT)在内的很多变换可以用于图像数据的变换。

表 8.11　有损 DPCM 均方根误差汇总

	劳埃德·马克斯量化器			自适应量化器		
预测器	2层	4层	8层	2层	4层	8层
式(8.5.16)	30.88	6.86	4.08	7.49	3.22	1.55
式(8.5.17)	14.59	6.94	4.09	7.53	2.49	1.12
式(8.5.18)	9.90	4.30	2.31	4.61	1.70	0.76

(续表)

预测器	劳埃德·马克斯量化器			自适应量化器		
	2 层	4 层	8 层	2 层	4 层	8 层
式(8.5.19)	38.18	9.25	3.36	11.46	2.56	1.14
压缩	8.00:1	4.00:1	2.70:1	7.11:1	3.77:1	2.56:1

图 8.28 显示了一个典型的变换编码系统。解码器执行的步骤(除了量化函数以外)与编码器是相反的。编码器执行 4 种相对简单的操作:子图分解、变换、量化和编码。一幅 $N\times N$ 大小的输入图像首先被分解为大小为 $n\times n$ 的子图像。这些子图进而被变换以生成 $(N/n)^2$ 个子图像变换阵列。每个阵列的大小为 $n\times n$。变换处理的目的是将每幅子图中的像素进行解相关,或用最少量的变换系数包含尽可能多的信息。在量化阶段有选择地消除或更粗略地量化带有最少信息的系数。这些系数对重构子图像质量的影响最小。编码处理在对量化的系数进行编码(通常使用变长码)后就终止了。任何或所有变换编码步骤都可以根据局部图像内容进行适应性调整,这被称为自适应变换编码。而如果这些步骤对所有子图都是固定的,则称为非自适应变换编码。

变换选择

基于各种二维离散变换的变换编码系统已经构建出来了或被广泛研究了。在给定的应用中选择特定的变换取决于可容忍的重构误差的大小和可用的计算资源。在变换系数的量化过程中(不是在变换阶段)实现了压缩。

考虑大小为 $N\times N$ 的图像 $f(x,y)$,该图像的正向离散变换 $T(u,v)$ 可以用通常的关系表示:

$$T(u,v)=\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}f(x,y)g(x,y,u,v) \qquad (8.5.24)$$

$u,v=0,1,2,\cdots,N-1$。给定 $T(u,v)$,$f(x,y)$ 可以同样使用一般的离散反变换计算得到:

$$f(x,y)=\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}T(u,v)h(x,y,u,v) \qquad (8.5.25)$$

$x,y=0,1,2,\cdots,N-1$。在这些公式中,$g(x,y,u,v)$ 和 $h(x,y,u,v)$ 分别称为正向和逆向变换核函数。由于在本节后面将会阐明的原因,这两个函数也被称为基础函数或基础图像。式(8.5.25)中的 $T(u,v)$,$u,v=0,1,2,\cdots,N-1$,称为变换系数;这些系数可以看做关于基础函数 $h(x,y,u,v)$ 的 $f(x,y)$ 一系列展开的展开系数(见 7.2.1 节)。

如果下列等式成立,则式(8.5.24)中的正向核函数称为是可分离的:

$$g(x,y,u,v)=g_1(x,u)g_2(y,v) \qquad (8.5.26)$$

另外,如果 g_1 在函数上等于 g_2,则核函数是对称的。此时,式(8.5.26)可以表示成下列形式:

$$g(x,y,u,v)=g_1(x,u)g_1(y,v) \qquad (8.5.27)$$



图 8.26 DPCM 结果图像。(a)1.0,(b)1.125,(c)2.0,(d)2.125,(e)3.0,(f)3.125 比特/像素

图 8.27 对应于图 8.26(a)到(f)的标定过的(×8)DPCM 误差图像

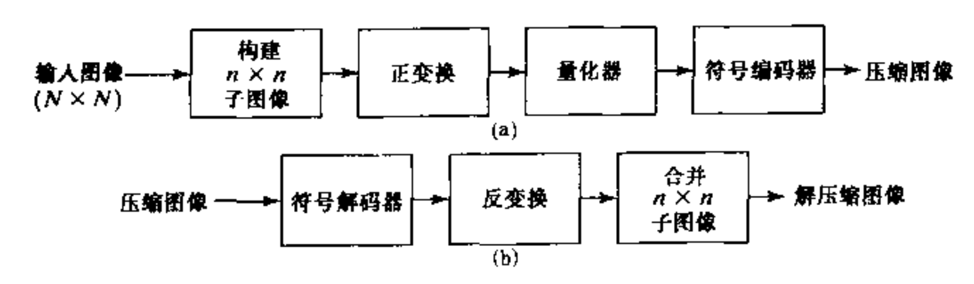


图 8.28 变换编码系统。(a)编码器,(h)解码器

如果 $g(x,y,u,v)$用式(8.5.26)和式(8.5.27)中的 $h(x,y,u,v)$代替,则对逆核也有相同的结论。说明具有可分离核的二维变换可以用对应的一维行-列或列-行变换计算得到,并不困难。4.6.1 节中解释了这种方法。

式(8.5.24)和式(8.5.25)中的正向和逆向变换核决定了所计算的变换类型和总体计算的复杂性以及所应用的变换编码系统的重构误差。最为有名的变换核函数对是:

$$g(x,y,u,v) = \frac{1}{N^2}e^{-j2\pi(ux+vy)/N} \tag{8.5.28}$$

和

$$h(x,y,u,v) = e^{j2\pi(ux+vy)/N} \tag{8.5.29}$$

这里 $j=\sqrt{-1}$。将这些核代入式(8.5.24)和式(8.5.25)得到在 4.2.2 节中介绍的离散傅里叶变换对的简化版本($M=N$)。

在变换编码过程中同样很有用处的一种在计算上更简单的变换称为 Walsh-Hadamard 变换(WHT)。这种变换是根据下列在函数上相同的核推导出来的:

$$g(x,y,u,v) = h(x,y,u,v) = \frac{1}{N}(-1)^{\sum_{i=0}^{m-1}[b_i(x)p_i(u)+b_i(y)p_i(v)]} \tag{8.5.30}$$

这里 $N=2^m$。求表达式的指数之和是以 2 为模的算术运算,并且 $b_k(z)$是 z 的二进制表达式的第 k 位(从右向左的)。例如,如果 $m=3$ 且 $z=6$(二进制值为 110),则 $b_0(z)=0$, $b_1(z)=1$, $b_2(z)=1$。式(8.5.30)中的 $p_i(u)$使用下面公式计算:

$$\begin{aligned} p_0(u) &= b_{m-1}(u) \\ p_1(u) &= b_{m-1}(u) + b_{m-2}(u) \\ p_2(u) &= b_{m-2}(u) + b_{m-3}(u) \\ &\vdots \\ p_{m-1}(u) &= b_1(u) + b_0(u) \end{aligned} \tag{8.5.31}$$

如前面所提到的,这里求和是以 2 为模的算术运算。对 $p_i(v)$有相似的表达式。

与 DFT 的核呈现正弦和余弦函数不同[见式(8.5.28)和式(8.5.29)],Walsh-Hadamard 核以棋盘形模式交替排列 +1 和 -1。图 8.29 显示了 $N=4$ 时的核。每个块包括 $4\times4=16$ 个元素(子方块)。白色表示 +1 而黑色表示 -1。为了得到左上角的块,令 $u=v=0$,并标出 $x,y=0,1,2,3$ 时 $g(x,y,0,0)$的值。此时所有的值均为 +1。顶部行中的第 2 个块是 $g(x,y,0,1)$在 $x,y=0,1,2,3$ 时取值的图,等等。正如已经看到的,Walsh-Hadamard 变换的重要意义在于它在实现上的简单性——所有核的值均为 +1 或 -1。

图像压缩中最常用到的一种变换是离散余弦函数变换(DCT)。这种变换是通过将下列(相等的)核代入式(8.5.24)和式(8.5.25)得到的:

$$\begin{aligned} g(x,y,u,v) &= h(x,y,u,v) \\ &= \alpha(u)\alpha(v)\cos\left[\frac{(2x+1)u\pi}{2N}\right]\cos\left[\frac{(2y+1)v\pi}{2N}\right] \end{aligned} \tag{8.5.32}$$

这里有：

$$\alpha(u)=\begin{cases}\sqrt{\dfrac{1}{N}} & u=0\\ \sqrt{\dfrac{2}{N}} & u=1,2,\cdots,N-1\end{cases} \tag{8.5.33}$$

对于 $\alpha(v)$ 的情况是类似的。图 8.30 显示了 $N=4$ 时的 $g(x,y,u,v)$。在计算方面遵循图 8.29中说明的格式，不同之处是 g 的值不是整数。在图 8.30 中，较浅的灰度级对应于更大的 g 值。

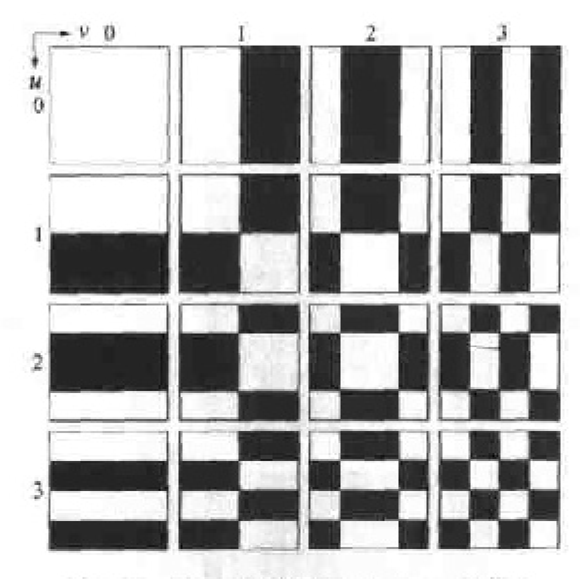


图 8.29 $N=4$ 时的 Walsh-Hadamard 基函数。每个块的原点在左上角

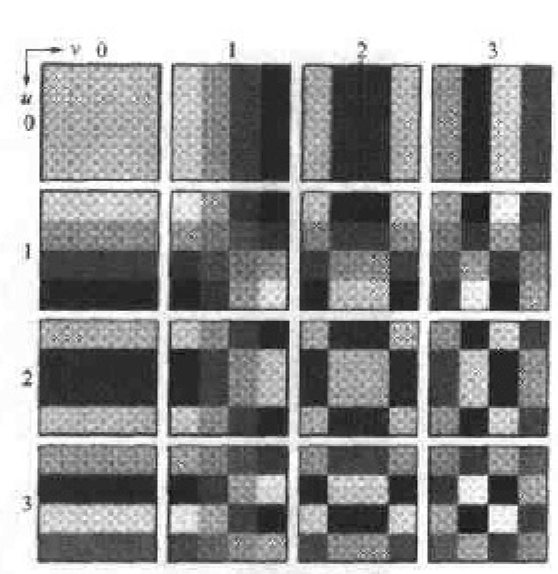


图 8.30 $N=4$ 时的离散余弦基函数。每个块的原点在块的左上角

例 8.19 分别使用 DFT，WHT 和 DCT 进行的变换编码

图 8.31(a)，(c)和(e)显示了对图 8.23 中的 512×512 大小的单色图像的 3 种近似。这些图像是这样得到的：先将原图分割为 8×8 大小的子图像，并使用刚才讲到的变换方法中的一种(即 DFT，WHT 或 DCT 变换)表示每一个子图像，将得到的系数的 50% 截去，再对截取的系数阵列进行逆变换。

在每种情况下，保留下来的 32 个系数是根据最大值的量级进行选择的。当不考虑任何量化或编码问题的时候，这个处理过程实际上使用一个等于 2 的因子对原图像进行压缩。注意在所有情况下，被丢弃的 32 个系数对重构图像品质的视觉影响是微乎其微的。然而，这些系数的消去伴随着产生了均方误差，这种均方误差的影响在图 8.31(b)，(d)和(f)的标定后的误差图像中可以见到。实际的 rms 误差分别为 1.28，0.86 和 0.68 个灰度级。

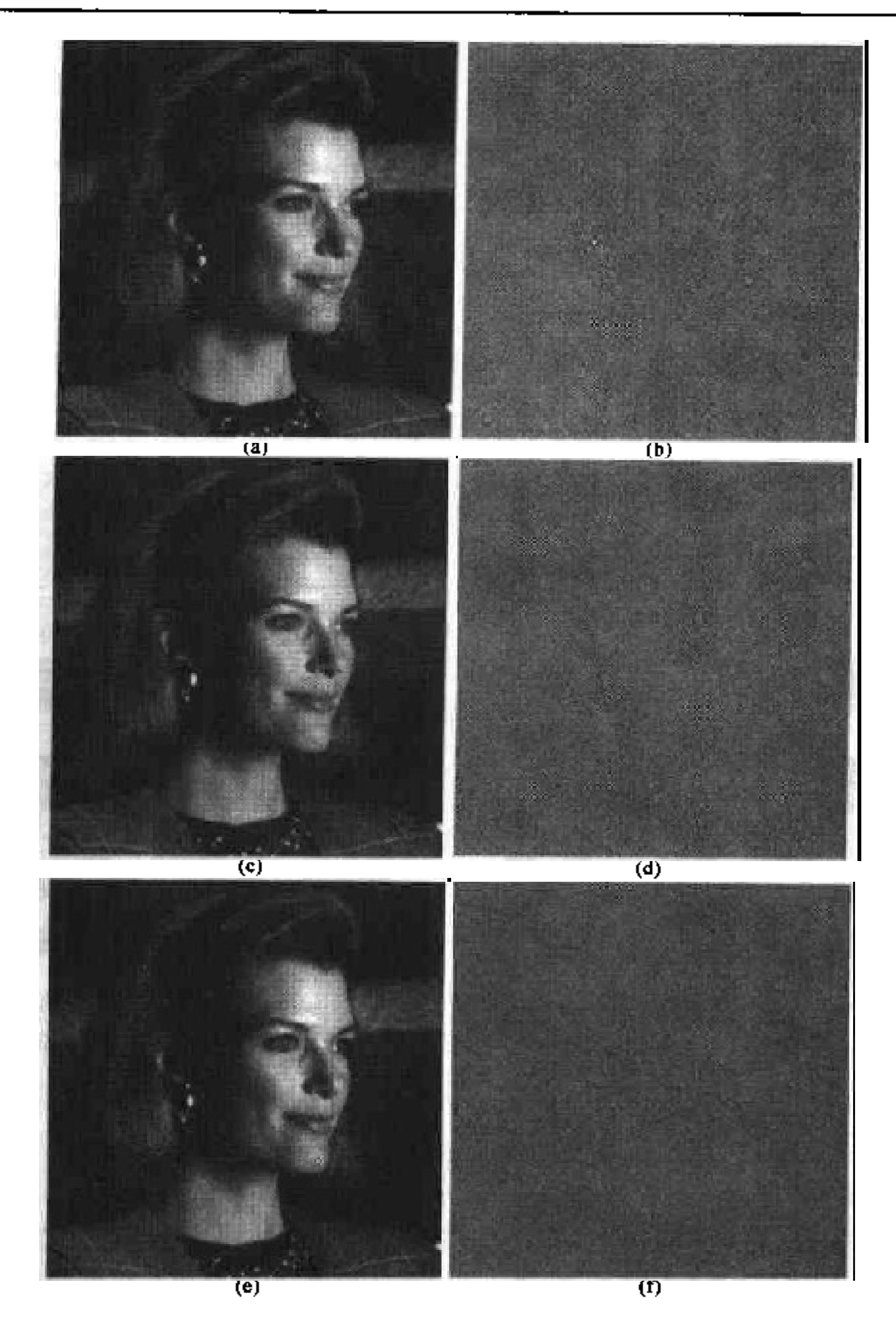


图 8.31 使用(a)傅里叶变换,(c)Hadamard 变换和(e)余弦变换对图8.23的近似,以及对应的标定后的误差图像

在前面的例子中注意到,均方重构误差中很小的差异与所使用的变换方法的信息包的特性或能量有直接关系。根据式(8.5.25),一幅 $n\times n$ 大小的图像$f(x,y)$可以表示为它的二维变换 $T(u,v)$的函数:

$$f(x,y)=\sum_{u=0}^{n-1}\sum_{v=0}^{n-1}T(u,v)h(x,y,u,v) \tag{8.5.34}$$

$x,y=0,1,\cdots,n-1$。注意,仅使用 n 代替了式(8.5.25)中的 N。现在考虑$f(x,y)$如何表示被压缩图像的子图。由于式(8.5.34)中的逆核 $h(x,y,u,v)$仅取决于参数 x,y,u,v,并且与

$f(x,y)$或 $T(u,v)$的值无关，所以，这个逆核函数可以看做式(8.5.34)定义的一系列基函数或基础图像集。如果对式(8.5.34)使用的符号进行如下一些修改，这种解释会变得更为清晰：

$$\mathbf{F} = \sum_{u=0}^{n-1}\sum_{v=0}^{n-1} T(u,v)\mathbf{H}_{uv} \tag{8.5.35}$$

这里 $\mathbf{F}$ 是一个包含 $f(x,y)$的像素的 $n \times n$ 矩阵，且有：

$$\mathbf{H}_{uv} = \begin{bmatrix} h(0,0,u,v) & h(0,1,u,v) & \cdots & h(0,n-1,u,v) \\ h(1,0,u,v) & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & & & \\ h(n-1,0,u,v) & h(n-1,1,u,v) & \cdots & h(n-1,n-1,u,v) \end{bmatrix} \tag{8.5.36}$$

而包含输入子图像素的矩阵 $\mathbf{F}$ 明确被定义为 n^2 个大小为 $n \times n$ 的矩阵的线性组合；它们是式(8.5.36)中的 $\mathbf{H}_{uv}$，$u,v=0,1,\cdots,n-1$。实际上，这些矩阵是式(8.5.35)的系列扩展的基础图像(或函数)；相关的 $T(u,v)$是扩展的系数。图 8.29 和图 8.30 用图形说明了 $n=4$ 时 WHT 和 DCT 的基础图像。

如果现在定义一个变换系数的模板函数：

$$\gamma(u,v) = \begin{cases} 0 & \text{如果 } T(u,v) \text{ 满足指定的截尾准则} \\ 1 & \text{其他} \end{cases} \tag{8.5.37}$$

$u,v=0,1,\cdots,n-1$，$\mathbf{F}$ 的一种近似可以根据截取后的展开得到：

$$\hat{\mathbf{F}} = \sum_{u=0}^{n-1}\sum_{v=0}^{n-1} \gamma(u,v)T(u,v)\mathbf{H}_{uv} \tag{8.5.38}$$

这里，构造 $\gamma(u,v)$消除了式(8.5.35)中总和的最小分布的基础图像。子图 $\mathbf{F}$ 和它的近似 $\hat{\mathbf{F}}$ 之间的均方误差为：

$$\begin{aligned} e_{ms} &= E\{\|\mathbf{F}-\hat{\mathbf{F}}\|^2\} \\ &= \mathbf{E}\left\{\left\|\sum_{u=0}^{n-1}\sum_{v=0}^{n-1} T(u,v)\mathbf{H}_{uv} - \sum_{u=0}^{n-1}\sum_{v=0}^{n-1}\gamma(u,v)T(u,v)\mathbf{H}_{uv}\right\|^2\right\} \\ &= E\left\{\left\|\sum_{u=0}^{n-1}\sum_{v=0}^{n-1} T(u,v)\mathbf{H}_{uv}[1-\gamma(u,v)]\right\|^2\right\} \\ &= \sum_{u=0}^{n-1}\sum_{v=0}^{n-1}\sigma^2_{T(u,v)}[1-\gamma(u,v)] \end{aligned} \tag{8.5.39}$$

这里，$\|\mathbf{F}-\hat{\mathbf{F}}\|$ 是矩阵$(\mathbf{F}-\hat{\mathbf{F}})$的范数，$\sigma^2_{T(u,v)}$是位置$(u,v)$的变换系数的方差。最终的化简是以基础图像的正交性质和假定 $\mathbf{F}$ 的像素是由零均值随机过程与已知的协方差产生的为基础进行的。因此，总的均方差近似误差是被丢弃的变换系数的方差之和；即，这些系数使 $\gamma(u,v)=0$，因此，式(8.5.39)中的$[1-\gamma(u,v)]$是 1。将最多的信息装入或重新分配到最少的系数中的变换提供了最好的子图近似，因此，也有最小的重构误差。最后，在导出式(8.5.39)的假设条件下，$(N/n)^2$ 个$N \times N$大小的图像的子图像的均方误差是相同的。因此，$N \times N$大小的图像的均方误差(是平均误差的一个量度)等于单个子图像的均方误差。

前面的例子显示出 DCT 的信息压缩能力比 DFT 和 WHT 的能力要强。尽管大多数自然图像通常都有这样的条件，但在数据压缩方面最佳的变换方法是 Karhunen-Loève 变换(第 11 章中对 KLT 变换进行了讨论)而不是 DCT。即，对任何输入图像和以任何数量保留下来的系数、

KLT 都可以将式(8.5.39)的均方误差降至最低(Kramer 和 Mathews[1956])[①]。然而,由于 KLT 具有数据依赖性,所以通常得到每幅子图的 KLT 基础图像所需的计算任务很繁重。由于这个原因,KLT 很少在实际的图像压缩中应用。反而像 DFT,WHT 或 DCT 这样的具有固定(有输入独立性)基础图像的变换方法的使用较为常见。对可能输入的独立变换,非正弦变换(如 WHT 变换)是最容易实现的。而正弦变换(如 DFT 或 DCT)的信息压缩能力更接近最佳 KLT 变换方法的能力。

因此,许多变换编码系统都是以 DCT 变换为基础的,这种变换方法在信息压缩能力和计算复杂性之间提供了一种很好的平衡。实际上,DCT 的性质已经提供了作为一种变换编码系统的国际标准所具有的实际价值(见 8.6 节)。对比其他输入独立的变换方法,DCT 变换具有使用单一的集成电路就可以实现,可以将最多的信息包装在最少的系数之中[②](对于大多数自然图像),以及能使被称为"分块噪声"的块效应最小等优点。这些分块噪声是由子图像之间的可见边界造成的。最后一条性质在与其他正弦变换进行比较时是非常重要的。如图 8.32(a)所示,DFT 所固有的 n 点周期性(见 4.6 节)造成了实质上的高频率变换内容的边界间断。当 DFT 变换系数被截尾或量化的时候,Gibbs 现象[③]是边界点呈现不正确的值,这些值在图像中以分块噪声的形式出现。即,相邻子图像之间的边界变得可见,因为子图像的边界像素具有在边界点构成间断的平均值[见图 8.32(a)]。图 8.32(b)的 DCT 变换减少了这种效应,因为这种变换方法隐含的固有 $2n$ 点周期不会产生固有的边界间断。

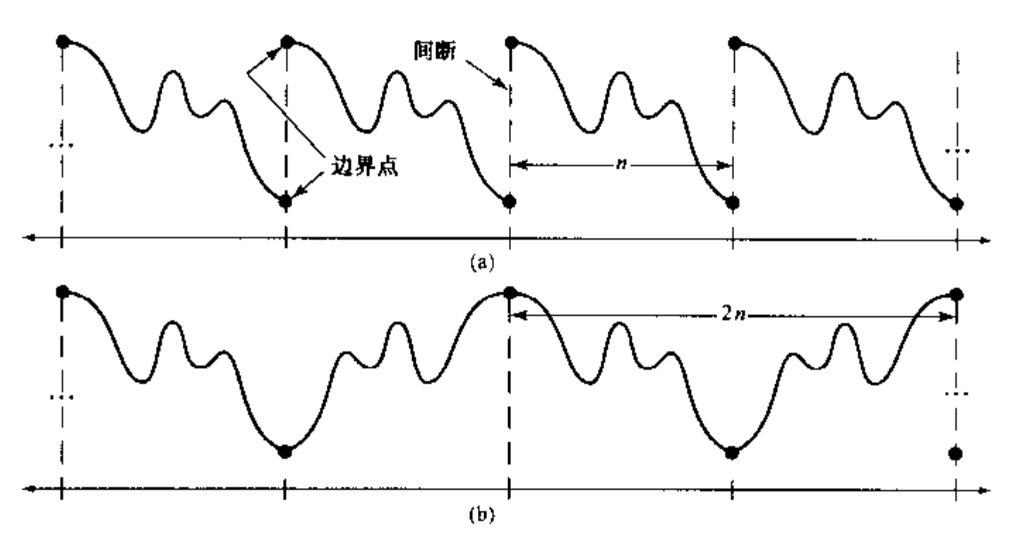


图 8.32 一维(a)DFT 和(b)DCT 的固有周期

① 最优性的另一个附加条件是式(8.5.37)的模板函数选择了最大值方差的 KLT 系数。

② Ahmed 等[1974]首先注意到一阶马尔可夫图像源的 KLT 基础图像与 DCT 基础图像非常相似。由于相邻像素之间的相关性接近 1,所以具有输入依赖性的 KLT 基础图像与具有输入独立性的 DCT 基础图像变成了同样的图像(Clarke[1985])。

③ 这种现象在大多数电路分析的电机工程内容中都有描述。它的产生是因为傅里叶变换无法在间断的地方均匀地收敛。在有间断的地方,傅里叶展开式取平均值。

子图像尺寸选择

另一个显著影响变换编码误差和计算复杂性的因素是子图像的大小。在大多数应用中，图像被进一步分割以便将相邻子图像之间的相关性(冗余)减少到可接受的程度，这里，n 是一个 2 的整数幂的数值，和前面一样，n 是子图的维度。后一个条件简化了子图像变换的计算(见 4.6.6 节中的逐次倍乘法)。通常，当子图像的尺寸增大的时候，压缩的水平和计算的复杂性也跟着增加。最常采用的子图像尺寸为 8×8 和 16×16。

例 8.20　变换编码中子图像尺寸的影响

图 8.33 图示说明了子图像的尺寸对变换编码重构误差的影响。图中所画的数据是通过将图 8.23 中的单色图像分割为大小为 $n \times n$ 的子图像(其中 $n = 2,4,8,16$ 和 32)，计算每幅子图像的变换，截取 75% 的系数，对截取后的阵列进行反变换得到的。注意，在子图像的尺寸大于 8×8 的时候，Hadamard 和余弦曲线会变得扁平，而反过来傅里叶重构误差在这个区域减小得更快。归纳这些曲线，n 值较大暗示着傅里叶重构误差将横穿 Walsh-Hadamard 曲线而收敛于余弦结果。实际上，这个结果与 Netravali 和 Limb[1980]以及 Pratt[1991]研究二维马尔可夫图像源时，在理论上提出的和实验上发现的结果相一致。

当使用的子图像大小为 2×2 的时候，所有的三条曲线都是相交的。此时，每种变换阵列只有四分之一的系数被保留。在所有情况下的系数都是直流分量，所以逆变换简单地使用它们的均值[见式(4.2.22)]对 4 幅子图的像素进行了替换。这种状况在图 8.34(d)中是很明显的，这幅图中显示了 2×2 大小的 DCT 变换结果放大的一部分。注意块效应，它在这个结果中是很普遍的，当子图像的大小增加到 4×4 和 8×8 的时候，图 8.34(e)和(f)中的块效应减少了。图 8.34(c)显示了作为参考的原图放大部分。另外，图 8.34(a)和(b)使这些结果与前面例子中的结果的比较变得容易了。

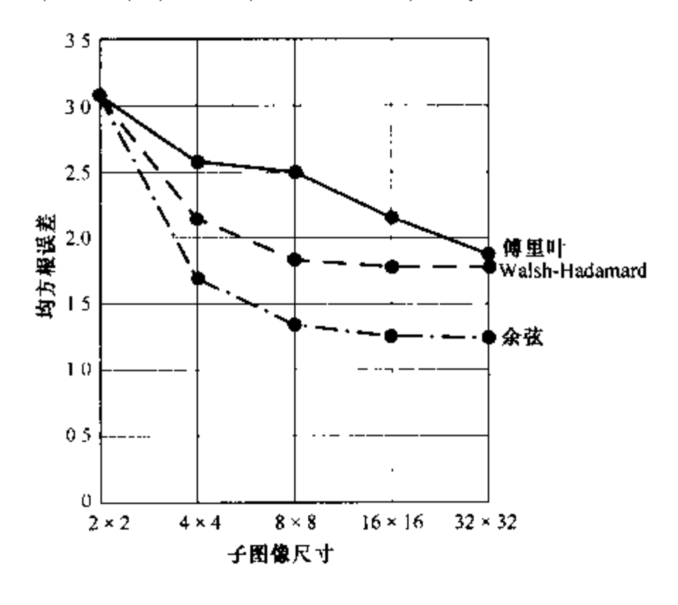


图 8.33　重构误差与子图像尺寸之比较

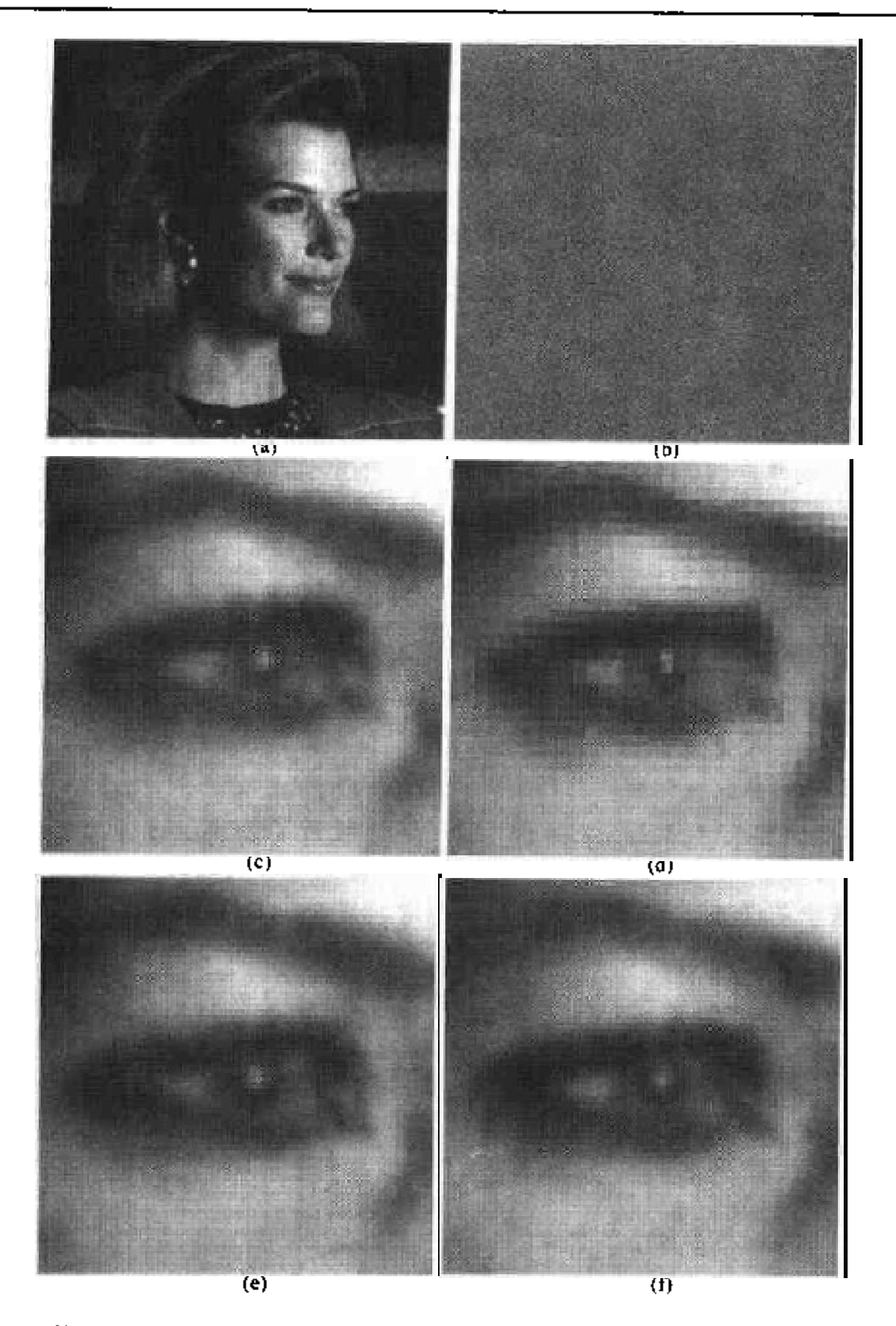


图 8.34 使用 25%的 DCT 系数对图 8.23 的近似。(a)和(b)8×8 大小的子图像结果,(c)放大的原图像,(d)2×2大小的结果,(e)4×4大小的结果,(f)8×8大小的结果

比特分配

与式(8.5.38)截取后的系列展开相关的重构误差,是一个被丢弃的变换系数的数量和相对重要性及用于表示保留系数的精度的函数。在大多数变换编码系统中,保留的系数是根据最大值方差进行选择的[即,建立式(8.5.37)的模板函数],称为区域编码,或根据最大值的量级选择,称为门限编码。对变换后的子图像的截取、量化和编码的整个过程通常称为比特分配。

例 8.21　比特分配

图 8.35(a)和(b)显示了对每个 8×8 子图像丢弃 87.5%的 DCT 系数的图 8.23 的两种近似。第一个结果是通过保留 8 个最大的变换系数的门限编码得到的。第二幅图像是使用区域编码方法生成的。在后一种情况下，每个 DCT 系数被看做一个随机变量，这个随机变量的分布可以通过对所有变换子图像的总体进行计算得到。8 个最大方差(占变换的 8×8 子图像中 64 个系数的 12.5%)的分布被定位并用它们判定系数 $T(u,v)$的坐标 u 和 v，这些系数将为所有的子图保留。注意，图 8.35(c)的门限编码差异图像包含的误差比图 8.35(d)中区域编码的误差少得多。图 8.35(e)和(f)提供了对(a)和(c)中重构图像一小部分的更接近的观察图像。

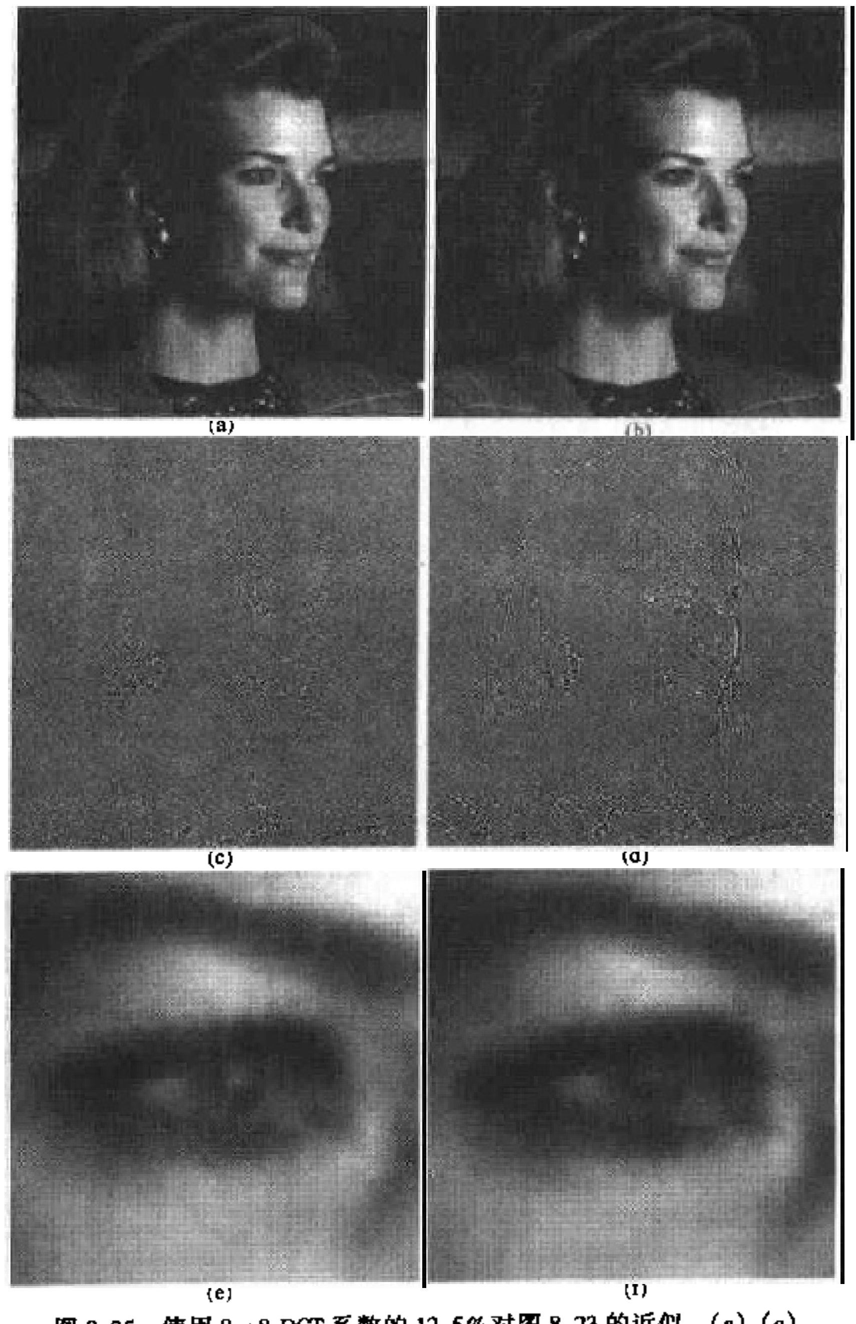


图 8.35　使用 8×8 DCT 系数的 12.5%对图 8.23 的近似。(a),(c)和(e)门限编码结果，(b),(d)和(f)区域编码结果

区域编码的实现

区域编码是以信息论中视信息为不确定性的概念为基础的。所以最大方差的变换系数携带着图像的大部分信息并在编码处理的过程中应该保留下来。这些方差本身可以就像前面的例子中那样直接根据总的$(N/n)^2$个变换子图像阵列计算出来,或以假设的图像模型(如,马尔可夫自相关函数)为基础进行计算。在任何一种情况下,根据式(8.5.38),区域取样处理可被看做每个$T(u,v)$用相应的区域模板中的元素相乘,其结果是在最大方差的位置上设为1,而在所有其他位置上置0构造出来的。最大方差的系数通常被定位在图像变换的原点周围。图8.36(a)中显示了典型的区域模板。

对在区域取样的过程中保留的系数必须进行量化和编码,这样,区域模板有时被描绘成显示用于对每个系数编码的比特数[图8.36(b)]。在大多数情况下,对系数分配相同的比特数,或在系数中不均匀地分配固定数目的比特数。在第一种情况下,系数通常用它们的标准差进行归一化并进行均匀量化。在第二种情况下,为每一个系数设计一个量化器,比如使用最佳劳埃德-马克斯量化器。为了构造所要求的量化器,通常使用瑞利密度函数对第零个或直流系数进行建模,而剩下的系数则用拉普拉斯或高斯密度[①]进行模拟。

分配给每个量化器的量化层次的数目(位数)与$\log_2\sigma_{T(u,v)}^2$成比例。这种分配是与率失真理论相一致的,这种理论表明方差为σ^2的高斯随机变量不能用少于$\frac{1}{2}\log_2(\sigma^2/D)$的比特来表示和用小于$D$的均方误差重建(见习题8.11)。直观的结论是高斯随机变量的信息内容与$\log_2(\sigma^2/D)$成比例。因此,式(8.5.38)中(当前讨论的上下文中)基于最大方差进行选择的保留系数可以被分配与系数方差的对数成比例的比特。

门限编码的实现

区域编码通常对所有的子图像使用单一固定的模板进行编码。而门限编码在为每幅彼此不同的子图像保留的变换系数的位置具有固有的自适应性。实际上,由于计算上的简单性,门限编码通常是实际中最常用的自适应变换编码方法。要强调的概念是,对任何子图像,最大量级的变换系数对重构子图像的品质具有最大的影响。这一点将在最后的例子中进行举例说明。因为不同子图像的最大系数的位置是变化的,所以通常$\gamma(u,v)T(u,v)$的元素重排列成一个一维行程编码序列。图8.36(c)对一幅假设图像的子图像显示了典型门限模板。这个模板提供了一种将对应子图像的门限编码处理过程可视化,及使用式(8.5.38)对过程进行数学描述的方便途径。当模板应用于[通过式(8.5.38)]产生这个模板的子图,并且根据图8.34(d)的z形排序模式,对得到的$n\times n$阵列进行重排序以组成一个n^2元素的系数序列的时候,记录下来的一维序列包含一些0的长行程[图8.36(d)中从0开始,按数字的顺序,z模式变得很明显]。这些过程通常是行程编码。对应于模板中一个1的位置的非零的或保留的系数用8.4节变长编码中的一种来表示。

① 由于每个系数都是子图中像素的线性组合[见式(8.5.24)],中心极限定理告诉我们:当子图尺寸增大的时候,系数就更趋向于高斯形式。但因为非负图像总有正的直流系数,所以这个结论不能应用于直流系数。

1	1	1	1	1	0	0	0
1	1	1	1	0	0	0	0
1	1	1	0	0	0	0	0
1	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

(a)

8	7	6	4	3	2	1	0
7	6	5	4	3	2	1	0
6	5	4	3	3	1	1	0
4	4	3	3	2	1	0	0
3	3	3	2	1	1	0	0
2	2	1	1	1	0	0	0
1	1	1	0	0	0	0	0
0	0	0	0	0	0	0	0

(b)

1	1	0	1	1	0	0	0
1	1	1	1	0	0	0	0
1	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

(c)

0	1	5	6	14	15	27	28
2	4	7	13	16	26	29	42
3	8	12	17	25	30	41	43
9	11	18	24	31	40	44	53
10	19	23	32	39	45	52	54
20	22	33	38	46	51	55	60
21	34	37	47	50	56	59	61
35	36	48	49	57	58	62	63

(d)

图 8.36　典型的(a)区域模板、(b)区域比特分配、(c)门限模板，和(d)门限系数排序序列。对保留的系数加深底纹

有 3 种基本途径对一幅变换后的子图像进行门限处理，或换种说法，有 3 种基本途径可以生成式(8.5.37)所给定形式的子图像门限模板函数：(1)对所有的子图像使用单一的全局门限；(2)对每幅子图像使用不同的门限；(3)门限可以随子图像中每个系数的位置函数的变化而变化。第 1 种方法中，对不同图像的压缩等级是不同的，这取决于超过全局门限系数的数目。第 2 种方法称为最大 N 编码，这种方法对每幅子图都丢弃相同数目的系数。结果是编码率是恒定的并且是事先可知的。第 3 种方法，类似于第 1 种方法，得到的编码率是变化的，但得到的好处是，可以通过使用下式代替式(8.5.38)中的 $\gamma(u,v)T(u,v)$ 实现门限处理和量化过程的结合：

$$\hat{T}(u,v) = \text{round}\left[\frac{T(u,v)}{Z(u,v)}\right] \tag{8.5.40}$$

这里 $\hat{T}(u,v)$ 是对 $T(u,v)$ 门限处理和量化的近似，$Z(u,v)$ 是变换的归一化矩阵：

$$\mathbf{Z} = \begin{bmatrix} Z(0,0) & Z(0,1) & \cdots & Z(0,n-1) \\ Z(1,0) & & & \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & & \cdot \\ Z(n-1,0) & Z(n-1,1) & \cdots & Z(n-1,n-1) \end{bmatrix} \tag{8.5.41}$$

在对子图像变换进行归一化(门限处理和量化)之前，可以对 $\hat{T}(u,v)$ 进行逆变换以得到子图

像$f(x,y)$的一个近似,$\hat{T}(u,v)$必须同 $Z(u,v)$相乘。得到的反向归一化阵列表示为 $\dot{T}(u,v)$,它是 $\hat{T}(u,v)$的近似:

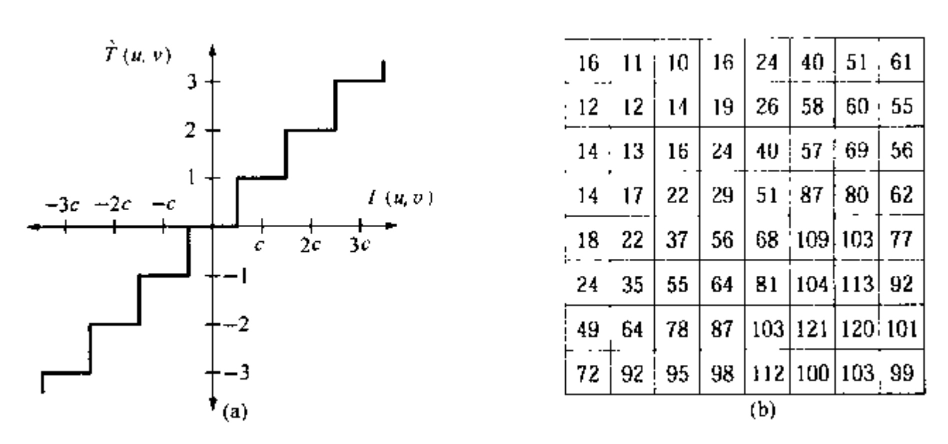
$$\dot{T}(u,v) = \hat{T}(u,v)Z(u,v) \tag{8.5.42}$$

$\dot{T}(u,v)$的逆变换生成解压缩子图像的近似。

图 8.37(a)用图形描绘了式(8.5.40)中 $Z(u,v)$被分配了一个特定的值 c 的情况。注意当且仅当下式成立的时候,$\hat{T}(u,v)$是整数值:

$$kc - \frac{c}{2} \leqslant T(u,v) < kc + \frac{c}{2}$$

如果 $Z(u,v) > 2T(u,v)$,则 $\hat{T}(u,v) = 0$ 且变换系数被完全截去或丢弃。当 $\hat{T}(u,v)$用一种变长编码来表示时,它随 k 的量级的增加,长度也增加,则用于表示 $T(u,v)$的比特数受到 c 值的控制。因此,**Z** 的元素可以成比例地在不同的压缩等级上缩放。图 8.37(b)显示了一种典型的标准化阵列。广泛地应用于 JPEG[①] 标准化中(见 8.6.2 节),这个阵列依照试探性的判定感觉或心理视觉的重要性对一幅变换过的子图像的每个系数进行加权。



16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

图 8.37 (a)一条门限编码量化曲线[见式(8.5.40)],(b)一个典型的标准化阵列

例 8.22 门限编码的说明

图 8.38(a)和(b)显示了两种对图 8.23 中的单色图像进行门限编码的近似结果。两幅图像都是使用一个 8×8 的 DCT 和图 8.37(b)的标准化阵列生成的。第 1 个结果的压缩率大约为 34:1,这是直接应用标准化阵列得到的结果。第 2 幅图以 67:1 的压缩率压缩原图,它是在用 4 乘以(比例缩放)标准化阵列后生成的。相比较之下,使用 8.4 节中讨论过的无误差方法得到的平均压缩率为 2.62:1。

图 8.23 中的原图和图 8.38(a)和(b)中的重构图像的差异分别显示在图 8.38(c)和(d)中。对应的 rms 误差[见式(8.1.8)]为 3.42 和 6.33 个灰度级。误差的准确特性在图 8.38(e)和(f)的放大图像中更为明显。这些图像分别显示了图 8.38(a)和(b)中的放大部分。这些图像对重构图像之间的微妙差异给予了更好的估算。

① JPEG 是联合图片专家组的简称。

(a) (b)

(c) (d)

(e) (f)

图 8.38　左列:对图 8.23 使用 DCT 和图 8.37(b)的标准化阵列处理得到的近似。右列:对4Z得到的相似结果

8.5.3 小波编码

类似前一节中介绍的变换编码系统,小波编码是基于以下概念:对一种将图像的像素解相关的变换系数进行编码比对原像素本身编码的效率更高。如果变换的基函数——此时为小波函数——将大多数重要的可视信息压缩到少量的系数中,则剩下的系数可以被粗略地量化或截取为0,而图像几乎没有失真。

图 8.39 显示了一个典型的小波编码系统。为了对 $2^J \times 2^J$ 大小的图像进行编码,选择并

使用一种分析小波 ψ 和最小分解层次 $J-P$ 对图像的离散小波变换进行计算。如果小波具有尺度函数,则可以用快速小波变换(见 7.4 节和 7.5 节)。在另一种情况下,计算出来的变换将原图像的一大部分转换为水平的、垂直的和对角的分解系数,这些系数具有零均值和类似拉普拉斯分布。回顾一下图 7.1 的图像和图 7.8(a)中图像的小波变换极简单的统计数字。由于许多计算的系数几乎不带有可视性的信息,这些系数的量化和编码可以实现最小的交互系数和编码冗余。另外,量化可以自适应地越过 P 分解层而利用任何位置关系。8.4 节中的一种或多种无损编码方法,包括行程长度、霍夫曼、算术方法和位平面编码,都可以应用到最终的符号编码步骤中。解码可以通过反向执行编码操作完成——除了量化过程以外,量化过程是不能逆向执行的。

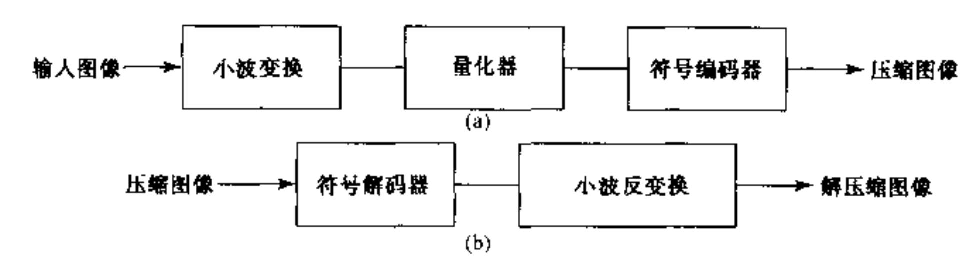


图 8.39 一个小波编码系统。(a)编码器,(b)解码器

图 8.39 以小波为基础的系统与图 8.28 的变换编码系统之间的主要差别是省略了变换编码器的子图像处理阶段。因为小波变换计算的有效性和固有的局部性(即,小波的基函数在宽度上是有限的),对原图进行细分是没有必要的。正如将在下面的例子中看到的,减少细分步骤可以消除分块噪声,这种噪声正是以 DCT 为基础的近似图像在高压缩比下的特点。

例 8.23 对比小波和基于 DCT 的编码

图 8.40 显示了图 8.23 中单色图像的两个基于小波的近似图像。图 8.40(a)是根据原图按 34:1 压缩的编码进行重构得到的图像。图 8.40(b)是按 67:1 压缩的编码进行重构得到的图像。由于这些比率对于例 8.22 的压缩等级是相同的,所以图 8.40(a)到(f)可以从量和质上与图 8.38(a)到(f)的基于变换的结果进行比较。

图 8.40(c)和(d)与图 8.38(c)和(d)分别进行可视性的比较,在小波编码结果中显示出误差的显著减少。实际上,图 8.40(a)中基于小波的图像的 rms 误差为 2.29 个灰度级,而图 8.38(a)中对应的基于变换的结果误差为 3.42 个灰度级。以相似的方法,图 8.38(b)和图 8.40(b)近似的 rms 误差分别为 6.33 和 2.96 个灰度级。在两个压缩等级上,计算出来的误差有利于基于小波的结果。

除了在给定的压缩等级上减少重构误差,小波编码[见图 8.40(e)和(f)]极大地增加(从主观方面讲)了图像的品质。在图 8.40(f)中这种改善特别明显。注意,在图8.38(f)中对应的基于变换的结果中占主要地位的分块噪声在这里没有出现。

当压缩水平超过 67:1 的时候(这是前面两个例子中分析得到的最大比率),在妇女衣服上的纹理丢失增加,她的眼睛也出现了模糊现象。这两种影响在图 8.41(a)和(b)中都是可见的,图8.41(a)和(b)是从图8.23 中原图像上进行的基于小波且压缩比率为 108:1 和167:1的编码进行重构得到的。增加的模糊效应在图 8.41(e)和(f)中特别明显。图 8.41(a)和(b)的 rms 误差分别为 3.72 和 4.73 个灰度级。对两幅图像进行主观评估的结果明显优于图 8.38(b)中

比率为67:1的基于变换得到的结果。它的rms误差为6.33个灰度级。因此,在压缩等级高出两倍多的情况下,多数高度压缩的基于小波的重构图像,仅有基于变换的、在较低压缩等级下得到的结果误差的75%,而且在感觉的品质上也优于后者。

为了总结关于小波压缩的讨论,我们对影响编码复杂性、性能和重构误差的主要因素进行了简要的总结。

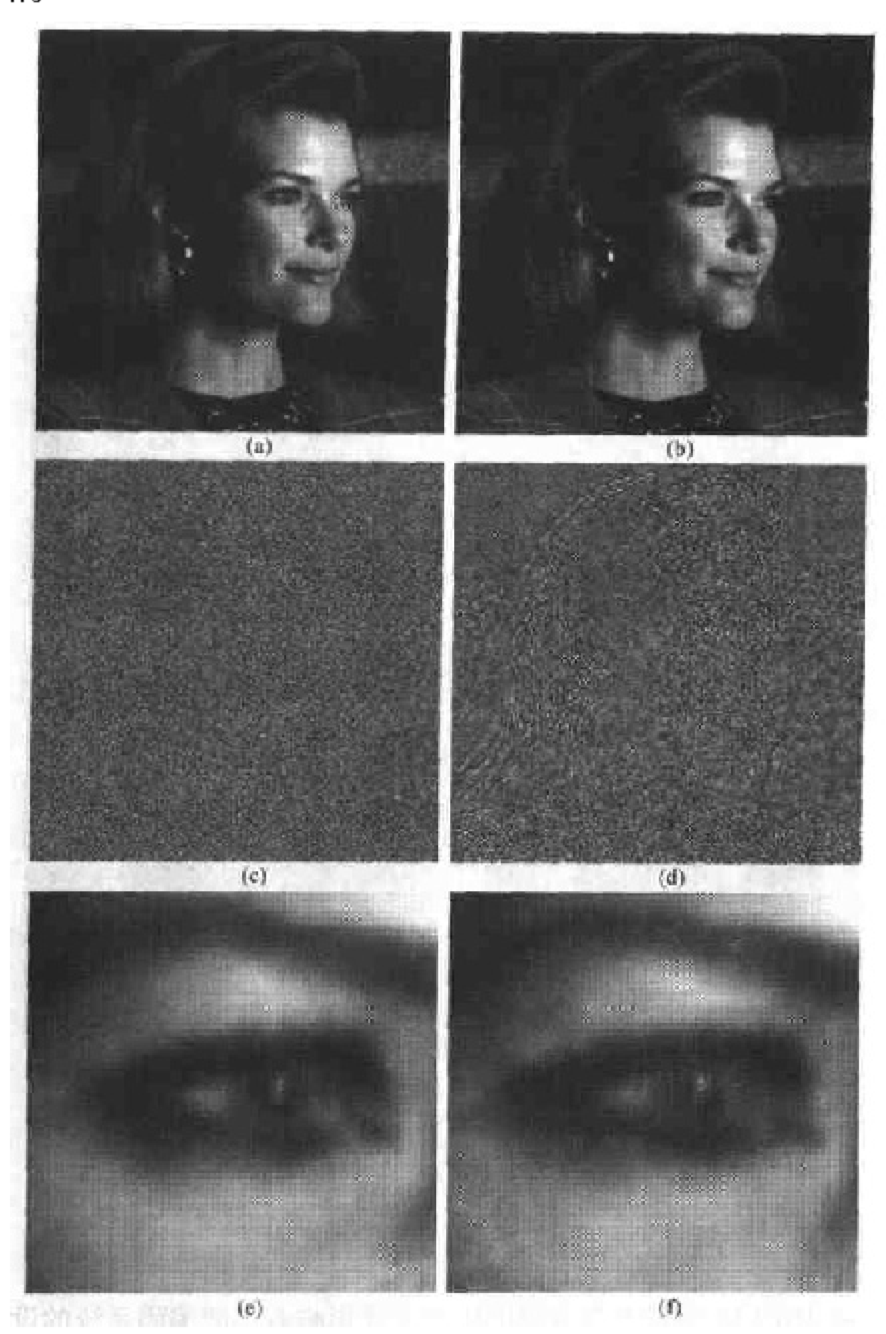


图8.40 (a),(c)和(e)小波编码的结果可以与图8.38(a),(c)和(e)基于变换的编码结果相比;(b),(d)和(f)与图8.38(b),(d)和(f)的结果相似

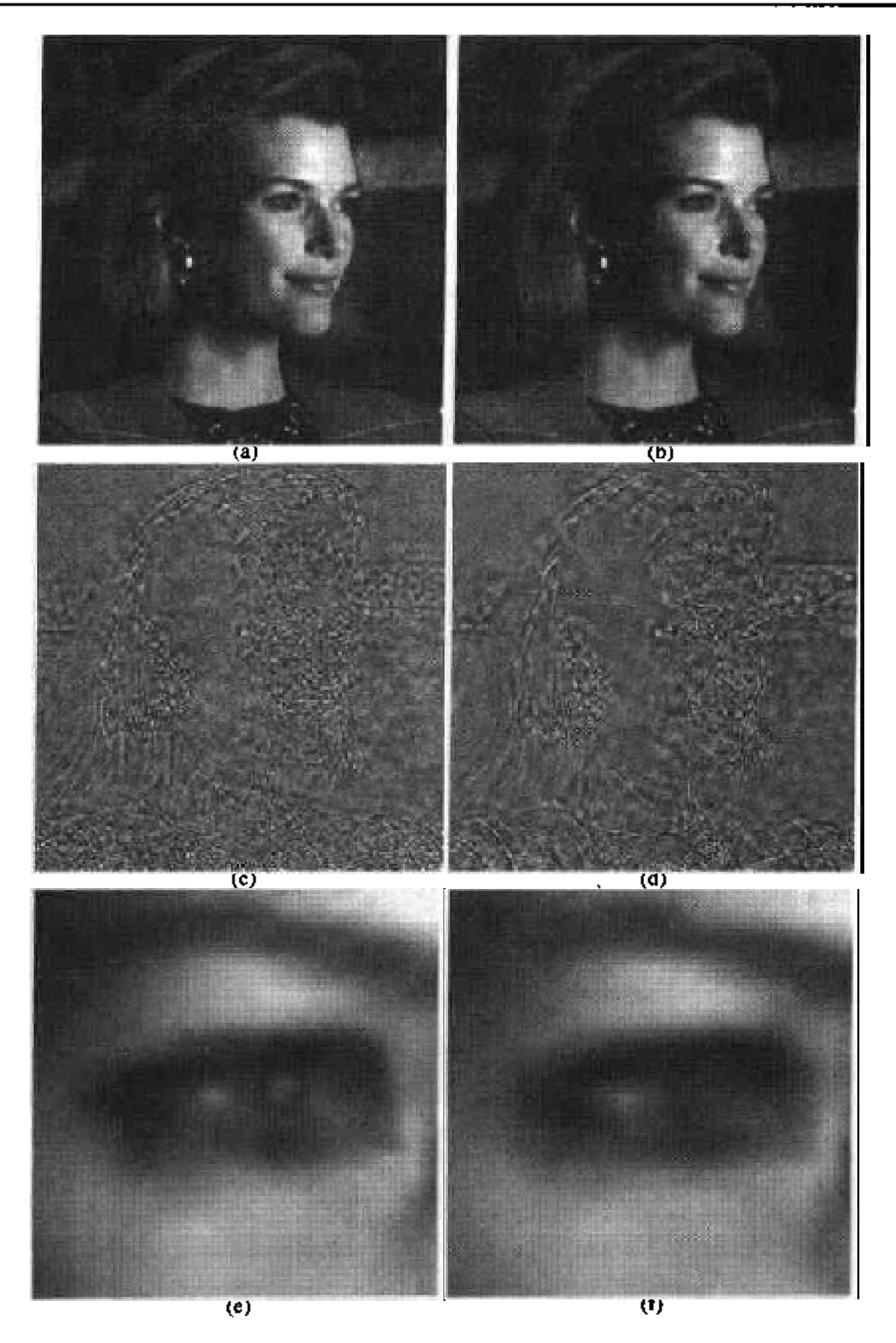


图 8.41 (a),(c)和(e)压缩率为 108:1 的小波编码结果,
(b),(d)和(f)压缩率为167:1的相似结果

小波选择

作为图 8.39 中正向和逆向变换基础的小波选择影响着小波编码系统的设计和性能的所有方面。它们直接影响着变换计算的复杂性和压缩及重构图像误差的可接受性。当变换小波带有一个尺度函数的时候,这种变换可以通过一系列数字滤波操作实现。滤波器抽头的数目等于非零小波和尺度向量系数的数目。小波将信息压缩为较少变换系数的能力取决于小波压缩和重构性能。

基于小波的压缩中最为广泛使用的展开函数是Daubechies小波和双正交的小波。后者允许将有用的分析特性融入分解滤波器，比如零矩的数目（见7.5节），而重要的综合特性、像重构的平滑度，则是融入重构滤波器。

例8.24 小波编码中的小波基

图8.42包含了在图8.23中的妇女图像的4个离散小波变换。哈尔（Haar）小波是最简单且是此处研究的惟一不连续的小波，这种小波用做图8.42(a)的展开或基函数。在图8.42(b)中使用了应用最普遍的成像小波中的Daubechies小波。而图8.42(c)中使用的对称小波（symlet）是具有增强的对称性的Daubechies小波的一种扩展。图8.42(d)所用的Cohen-Daubechies-Feauveau小波用于说明双正交小波的性能。正如以前的这类例子一样，所有细节的系数都进行了标定以便能使重点强调的结构更为突出——灰度级128对应系数值为0。

正如在表8.12中看到的，从图8.42(a)到(d)移动的过程中，图8.42中的变换所包含的对每个系数（在每个分解的等级上）进行乘法和加法计算的操作次数从4次增加到28次。所有4种变换都使用一种快速小波变换（也就是滤波器组）公式进行计算。注意，当计算的复杂性（也就是滤波器抽头的数目）增加的时候，信息压缩能力会更好。当使用哈尔小波和低于1.5的细节系数被舍入为0的时候，总变换系数的46%均被清零。对于更复杂的双正交小波，被清零的系数数目增加到55%，增加的潜在压缩量近10%。

表8.12 当对图8.42中变换的截取低于1.5的时候，小波变换滤波器抽头和调零系数

小波	滤波器抽头（尺度+小波）	调零系数
哈尔小波（见习题7.10）	2+2	46%
Daubechies小波（见图7.6）	8+8	51%
Symlet小波（见图7.24）	8+8	51%
双正交小波（见图7.37）	17+11	55%

分解等级选择

另一种影响小波编码计算复杂性和重构误差的因素是变换分解等级的选择。由于P尺度快速小波变换涉及P个滤波器组迭代，正向和反向变换计算操作的次数随着分解级别的提高而增加。再有，对逐渐增加的低尺度系数的量化导致更多的分解级别，影响着重构图像的逐渐增大的区域。在很多应用中（像搜索图像数据库，或对逐渐重构传送图像等），存储或传送图像的分辨率和最低可用近似图像的尺度通常决定了变换级别数。

例8.25 小波编码中的分解级别

表8.13说明了根据图8.23中固定全局门限值为25的编码选择分解等级的影响。和以前的小波编码例子一样，只对细节系数进行舍入。表中列出了清零的系数的百分比和得到的rms重构误差。注意，初始时的分解是大多数数据压缩的原因。在3个分解等级上舍入的系数在数字上几乎没有差别。

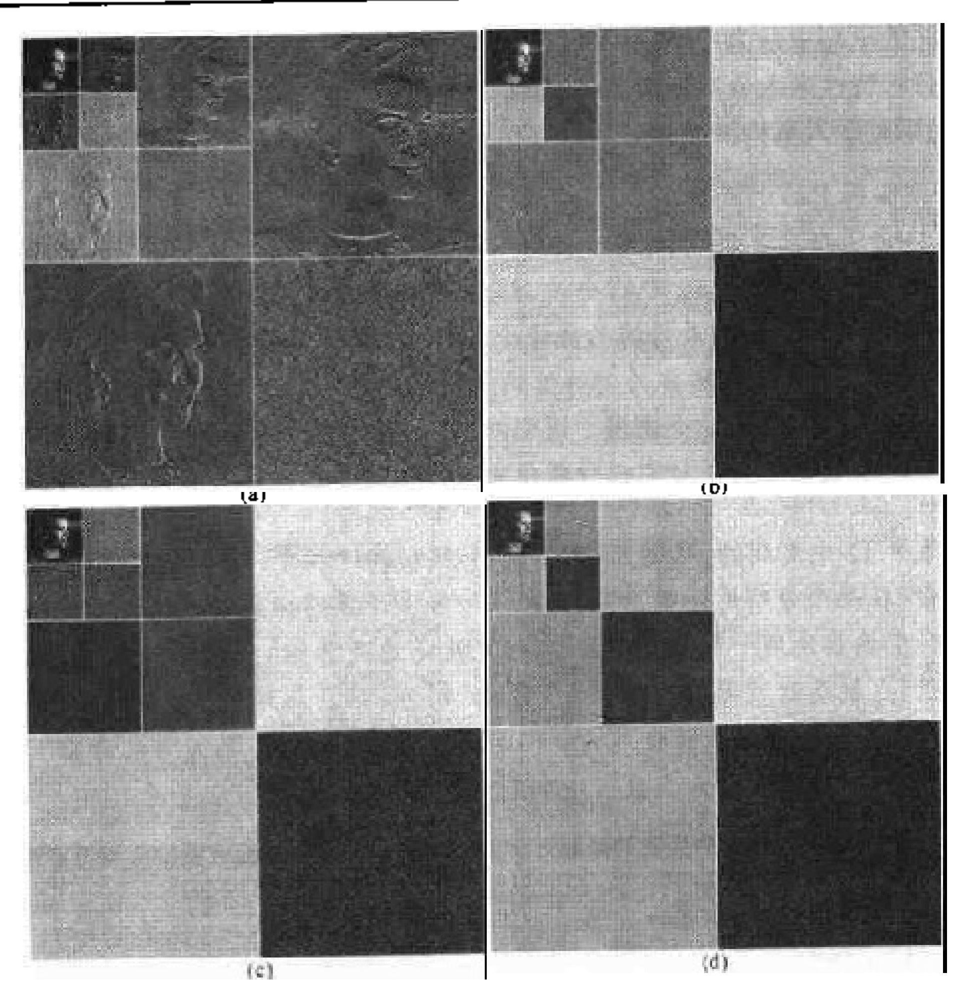


图 8.42 图 8.23 的小波变换。(a)使用哈尔小波,(b)使用 Daubechies 小波,(c)使用对称小波和(d)Cohen-Daubechies-Feauveau 双正交小波

表 8.13 对图 8.23 中 512×512 大小的图像进行小波编码时分解级别的影响

尺度和滤波器组迭代	近似系数图像	舍入的系数(%)	重构误差(rms)
1	256×256	75%	1.93
2	128×128	93%	2.69
3	64×64	97%	3.12
4	32×32	98%	3.25
5	16×16	98%	3.27

量化设计

影响小波编码压缩和重构误差的最大因素是系数量化。尽管最广泛应用的量化器是均匀的,但量化的效果可以通过以下方法进一步改进:(1)引入一个以零为中心的扩大的量化间隔,这个间隔被称为死区,或(2)从一个尺度到另一个尺度自适应调整量化间隔的大小。不论哪种情况,选择的量化间隔都必须随着编码图像的比特流传送给解码器。这个间隔本身根据压缩的图像,通过不断的试探或自动的计算得到。例如,一个全局系数门限可以作为第一级细节系

数绝对值的中值,或作为舍去的零的数目和保留在重构图像中能量总量的函数计算得到。

例 8.26　在小波编码中的死区间隔选择

图 8.43 说明了与截尾的细节系数的百分比相关的死区间隔大小对图 8.23 中的妇女图像基于三尺度双正交小波编码的影响。当死区尺寸增大的时候,截尾系数的数目同样增大。在曲线拐点处的上方(即,横坐标 4.5 处),几乎没有增益。这是由于细节系数直方图在零的周围有很高的尖峰(见图 7.8 的例子)。

与图 8.43 中死区门限对应的 rms 重构误差在门限 4.5 处从 0 增加到 1.77 个灰度级,并在门限为 18 的地方增加到 2.79 个灰度级。在门限为 18 的地方的零系数数目达到了 96.43%。如果每个细节系数都被消掉,这个百分比将上升 1.5%,但重构误差会增加到 7.6 个灰度级。

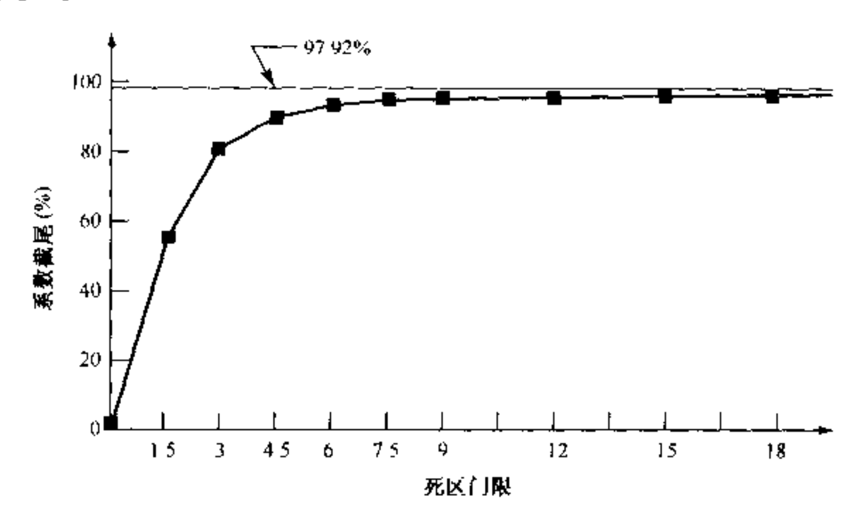


图 8.43　关于小波编码的死区区间选择的影响

8.6　图像压缩标准

迄今为止讨论的许多有损和无误差压缩方法在广泛使用的图像压缩标准中起着重要作用。在本小节中,对其中的几种标准进行一些分析,并使用这些标准说明一下前面提到的压缩方法。讨论的大多数标准都是得到国际化标准组织(ISO)和国际电话与电报咨询委员会(CCITT)认可的。这些标准适用于二值图像和连续色调(单色的和彩色的)图像的压缩,同时也适用于静止画面和视频图像(即,连续帧)。

8.6.1　二值图像压缩标准

两种应用最为广泛的图像压缩标准是用于二值图像压缩的 CCITT 第三组和第四组标准。尽管它们现在应用于多种计算机应用领域,但它们最初是为了使用传真(FAX)编码方法通过电话网络传输文件而设计的。第三组标准应用非自适应的 1 维行程编码技术,在这种技术中,对每组 K 条线($K=2$ 或 4)的最后 $K-1$ 条线用二维方式进行可任意选择的编码。第四组标准是第三组标准进行简化和流水线化得到的版本,在该标准中,只允许存在二维编码。两组标准使用同样的非自适应的二维编码方法。这种方法与 8.4.2 节中描述的相对地址编码(RAC)

技术非常相似。

在建立 CCITT 标准的过程中,有八种典型的“测试”文件被选择作为评估不同的二值压缩方案的基准。用现存的第三组和第四组标准对这些包括打印和手写内容(使用几种不同的语言)以及少量线条的文件进行比率约为 15:1 的压缩。然而,由于第三组和第四组标准是以非自适应技术为基础的,所以有时候会导致数据膨胀(如对半色调图像即是如此)。为了克服这种现象以及出现的相关问题,联合二值图像专家组(JBIG)——这是一个 CCITT 和 ISO 的联合委员会——已经采用和/或建议了几种其他的二值压缩标准。这其中包括 JBIG1 标准,一种自适应算术压缩技术,这是目前可用的处理最一般情况和最坏情况下的二值压缩的技术,还有 JBIG2 标准(现在还是一份委员会的最终草案)。使用这种标准得到的压缩效果通常是使用 JBIG1 标准得到的效果的 2 到 4 倍。这些标准可以同时应用于二值图像和高达 6 灰度编码比特/像素的灰度级图像(以一个位平面为基础)的压缩。

一维压缩

在一维 CCITT 第三组压缩方法中,图像[①] 的每一条线都可以用一系列变长编码码字编码,这些码字代表从左到右扫描线条过程中,白色和黑色交替的行程长度。码字本身分两类。如果行程长度小于 63,则使用表 8.14 中修正的霍夫曼编码得到的一个终结编码。如果行程长度大于 63,则根据表 8.15 得到最大可能出现的组成编码(不超过行程长度),将它与一个终结编码一起使用进行编码,终结编码用于表示组成编码和实际行程长度之间的差异。这个标准要求每条线都从一个白色行程长度码字开始,事实上它们可能是 00110101,这个编码表示一个零长度的白色行程。最后,惟一的行尾(EOL)码字 000000000001 用于结束每一行,同时标记每幅新图像的第一行。一个图像序列的结尾使用 6 个连续的 EOL 标记。

表 8.14 CCITT 终结编码

扫描宽度	白色码字	黑色码字	行程长度	白色码字	黑色码字
0	00110101	0000110111	32	00011011	000001101010
1	000111	010	33	00010010	000001101011
2	0111	11	34	00010011	000011010010
3	1000	10	35	00010100	000011010011
4	1011	011	36	00010101	000011010100
5	1100	0011	37	00010110	000011010101
6	1110	0010	38	00010111	000011010110
7	1111	00011	39	00101000	000011010111
8	10011	000101	40	00101001	000001101100
9	10100	000100	41	00101010	000001101101
10	00111	0000100	42	00101011	000011011010
11	01000	0000101	43	00101100	000011011011
12	001000	0000111	44	00101101	000001010100
13	000011	00000100	45	00000100	000001010101

① 在标准中,图像被看做是页,而图像序列被看做文件。

（续表）

扫描宽度	白色码字	黑色码字	行程长度	白色码字	黑色码字
14	110100	00000111	46	00000101	000001010110
15	110101	000011000	47	00001010	000001010111
16	101010	0000010111	48	00001011	000001100100
17	101011	0000011000	49	01010010	000001100101
18	0100111	0000001000	50	01010011	000001010010
19	0001100	00001100111	51	01010100	000001010011
20	0001000	00001101000	52	01010101	000000100100
21	0010111	00001101100	53	00100100	000000110111
22	0000011	00000110111	54	00100101	000000111000
23	0000100	00000101000	55	01011000	000000100111
24	0101000	00000010111	56	01011001	000000101000
25	0101011	00000011000	57	01011010	000001011000
26	0010011	000011001010	58	01011011	000001011001
27	0100100	000011001011	59	01001010	000000101011
28	0011000	000011001100	60	01001011	000000101100
29	00000010	000011001101	61	00110010	000001011010
30	00000011	000001101000	62	00110011	000001100110
31	00011010	[illegible]			

[illegible]

(续表)

	码 字		码 字
1792	00000001000	2240	000000010110
1856	00000001100	2304	000000010111
1920	00000001101	2368	000000011100
1984	000000010010	2432	000000011101
2048	000000010011	2496	000000011110
2112	000000010100	2560	000000011111
2176	000000010101		

二维压缩

为 CCITT 第三组和第四组标准所采用的二维压缩方法采用的是逐线方法,这种方法在每个黑色转白色或白色转黑色的扫描转换位置上均参考基准元素 a_0 进行编码,基准元素 a_0 被设定在当前的编码线上。前面提到的编码线称为基准线;对每幅新图像的第一条线设定的基准线是一条虚构的白色线条。

图 8.44 显示了对一条单扫描线的基本编码过程。注意,这个过程的初始步骤在于对几个关键的转换或变化元素的定位:a_0,a_1,a_2,b_1 和 b_2。变化元素定义为在同一条直线上与前一个像素值不同的像素。最重要的变化元素是 a_0(基准元素),这个元素被设定在虚构的白色变化元素的位置上,而这个虚构变化元素的位置在每条新编码线的第一个像素的左边,或者 a_0 可以根据以前的编码模式确定。在 a_0 的位置确定了之后,a_1 作为在当前编码线上 a_0 右边的下一个变化元素的位置,a_2 作为在当前编码线上 a_1 右边的下一个变化元素的位置,b_1 作为具有(a_0 的)相反值的变化元素,位于基准线(或前一条线)上 a_0 的右边。b_2 作为下一个变化元素位于基准线上 b_1 的右边。如果这些变化元素中的任何一个没有被检测到,则这些元素被设定在适当线上最后一个像素右边的一个虚构像素位置上。图 8.45 给出了在不同变化元素之间一般关系的两种说明。

在对当前基准元素和与之相联系的变化元素进行辨别之后,就要执行两种测试,以便从三种可能的编码模式中选择一种。这三种编码模式分别为:通过模式、垂直模式和水平模式。对应于图 8.44 中流程图的第一分支节点的初始测试比较 b_2 相对 a_1 的位置。对应于图 8.44 中第二个分支点的第二个测试是计算 a_1 和 b_1 位置之间的距离,并将此距离与 3 进行比较。依靠这些检测的输出,将进入图 8.44 中的三个编码框图中的一个,并执行相应的编码程序。然后,设置一个新的基准元素,如流程图中画的一样,开始准备下一次重复编码。

表 8.16 定义了被三种可能的编码模式用到的特定编码。在通过模式中很明显地排除了 b_2 直接位于 a_1 上方的情况,而只需要通过模式的码字 0001。如图 8.45(a)所示,这种模式识别与当前的白色或黑色编码线的扫描没有重叠的白色或黑色基准线扫描段。在水平编码模式中,从 a_0 到 a_1 和从 a_1 到 a_2 的距离必须按与表 8.14 和表 8.15 中的终结和组成编码相关的方式编码,并且要附加到水平模式码字 001 上。表 8.16 中由符号 $001 + M(a_0a_1) + M(a_1a_2)$ 对这种模式进行了说明。这里 a_0a_1 和 a_1a_2 分别代表从 a_0 到 a_1 和从 a_1 到 a_2 的距离。最

后，在垂直模式中，六个特定的变长编码之一被分配给 a_1 和 b_1 之间的距离。图 8.45(b)说明了在水平和垂直模式编码中用到的参数。在表 8.16 的底部，扩展模式码字用于选取一种可选的传真编码模式。例如，编码 0000001111 用于初始化一种传输的非压缩模式。

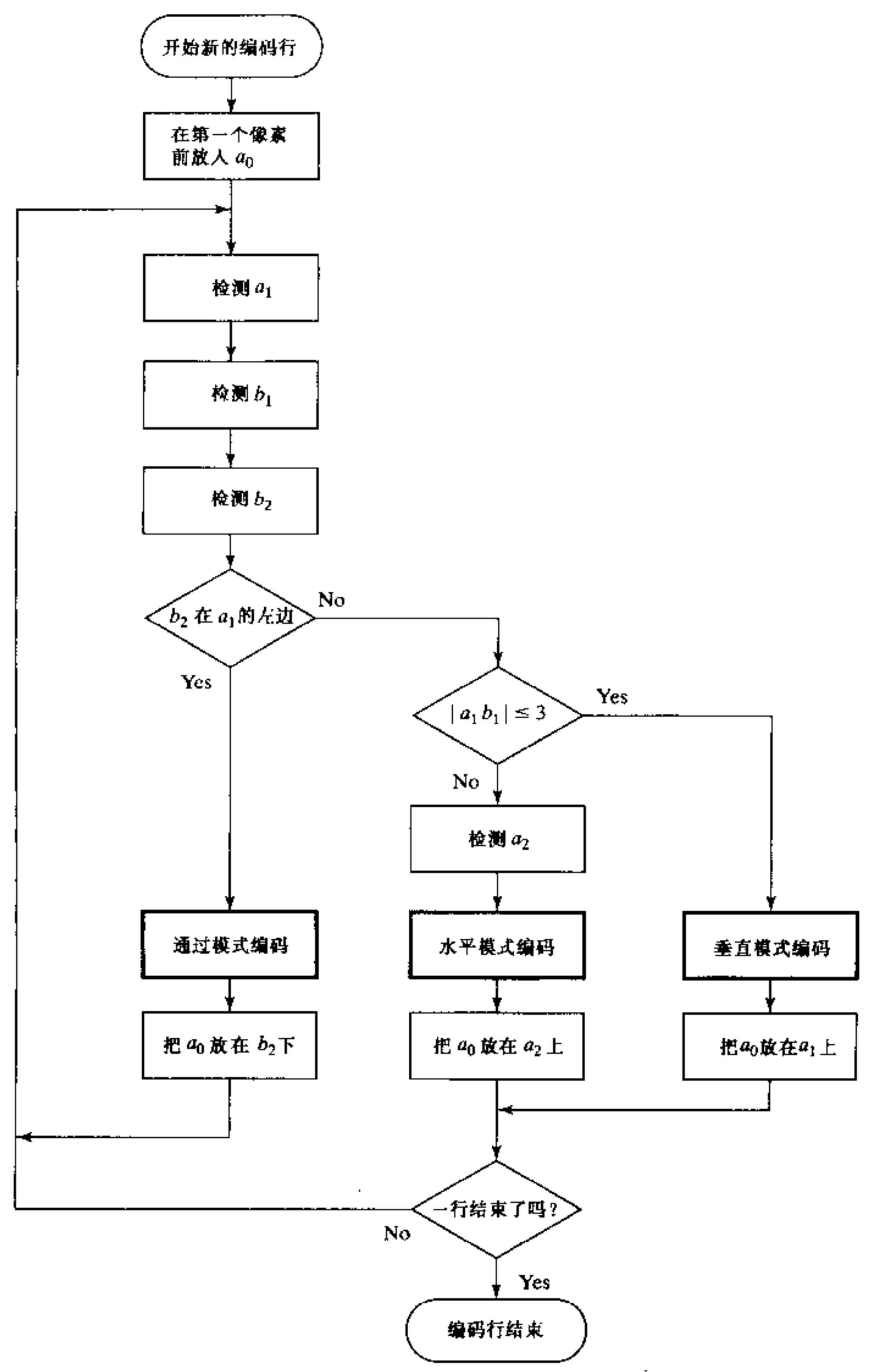


图 8.44　CCITT 二维编码流程。符号 $|a_1 b_1|$ 代表变化元素 a_1 和 b_1 之间距离的绝对值

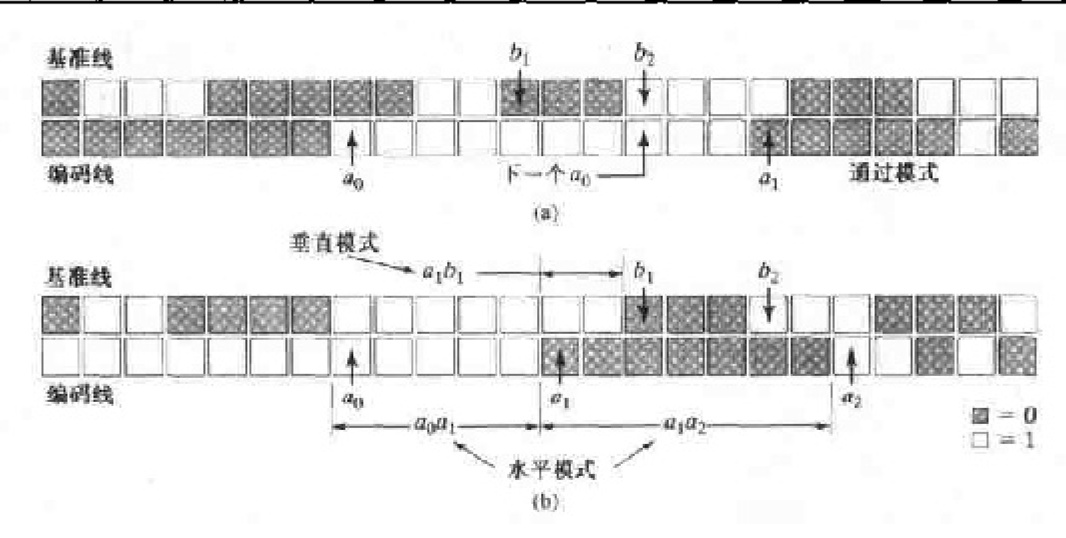


图 8.45 CCITT(a)通过模式和(b)水平模式与垂直模式编码参数

表 8.16 CCITT 二维编码表

模 式	码 字
通过	0001
水平	$001+M(a_0a_1)+M(a_1a_2)$
垂直	
a_1 在 b_1 之下	1
a_1 在 b_1 右边相距一个像素的地方	011
a_1 在 b_1 右边相距两个像素的地方	000011
a_1 在 b_1 右边相距三个像素的地方	0000011
a_1 在 b_1 左边相距一个像素的地方	010
a_1 在 b_1 左边相距两个像素的地方	000010
a_1 在 b_1 左边相距三个像素的地方	0000010
扩展	0000001×××

例 8.27 CCITT 垂直模式编码实例

尽管图 8.45(b)对水平和垂直模式编码都有参数注释,但实际上描绘的情况是垂直模式编码的情况。即,当 b_2 在 a_1 的右边时,图 8.44 中的第一种(或通过模式)检测无法通过。第二个测试决定是垂直还是水平编码模式进入。这个检测表明应该使用垂直模式编码,因为从 a_1 到 b_1 的距离小于 3。根据表 8.16,正确的码字为 000010,这意味着 a_1 在 b_1 左边相距两个像素的位置上。在下一步重复编码的准备过程中,a_0 移动到 a_1 的位置上。

8.6.2 连续色调静止图像压缩标准

CCITT 和 ISO 已经定义了几种连续色调(与二值相对应)图像压缩标准。这些在不同程度上被认可的标准都是用于处理单色和彩色图像压缩的标准。与 8.6.1 节中描述的二值压缩标准相反,连续色调标准主要以 8.5.2 节和 8.5.3 节的有损变换编码技术为基础。为了进一步研制这些标准,CCITT 和 ISO 委员会向很多公司、大学和研究实验室征求算法建议。根据图像

的品质和压缩的效果从提交的方案中选择最好的算法。这样得到的标准展示了在连续色调图像压缩领域的现有水平。这其中也包括原来的基于 DCT 的 JPEG 标准,最近提出的基于小波的 JPEG 2000 标准以及 JPEG-LS 标准。JPEG-LS 标准是一种无损到接近无损的自适应预测方案,它包括对平面区域检测和行程编码(ISO/IEC[1999])的机理。

JPEG

使用最为普遍且易于理解的连续色调静止帧压缩标准是 JPEG 标准。这种标准定义了三种不同的编码系统:(1)一种有损基本编码系统,这个系统是以 DCT 为基础的并且足够应付大多数压缩方面的应用;(2)一种扩展的编码系统,这种系统面向的是更大规模的压缩,更高的精确性或逐渐递增的重构应用系统;以及(3)一种面向可逆压缩的无损独立编码系统。为了实现 JPEG 的兼容性,产品或系统必须包含对基本系统的支持。没有规定特殊的文件格式、空间分辨率或彩色空间模型。

在经常被称为连续基准系统的基准系统中,输入和输出数据的精度限制为 8 比特,而量化的 DCT 值限制为 11 比特。压缩过程本身包括三个连续的步骤:DCT 计算,量化以及变长编码分配。图像首先被细分为 8×8 的像素块。这些像素块进行从左到右,从上到下的处理。当对每个 8×8 的块或子图像进行了处理之后,通过减去 2^{n-1} 对 64 个像素进行层次移动,2^n 是灰度级的最大数目。然后计算块的二维离散余弦变换,根据式(8.5.40)进行量化并重排。使用图 8.36(d)的 z 形模式形成一个量化系数的一维序列。

由于在图 8.36(d)中的 z 形模式下生成的一维重排阵列是根据递增的空间频率定性地进行排列的,所以 JPEG 编码程序的设计可以充分利用根据重排得出的零的长扫描段优点。特别是非零的 AC[①] 系数是使用规定了系数值和处在前面位置的零的个数的一种变长编码进行编码的。DC 系数相对于前面子图像的 DC 系数进行不同的编码。表 8.17、表 8.18 和表 8.19 给出了对亮度图像默认的 JPEG 霍夫曼编码。JPEG 推荐的亮度量化阵列在图 8.37(b)中给出,并且可以依比例的不同而提供多种压缩等级。尽管对亮度和色度处理提供了默认的编码表和证明过的量化阵列,但用户可以自由地构造自定义的表格和/或阵列,这样可以更好地适应被压缩图像的特点。

例 8.28 JPEG 基准编码和解码

考虑下列 8×8 子图像,使用 JPEG 基本标准进行压缩和重构:

52	55	61	66	70	61	64	73
63	59	66	90	109	85	69	72
62	59	68	113	144	104	66	73
63	58	71	122	154	106	70	69
67	61	68	104	126	88	68	70

① 在标准中,AC 项表示除了零或 DC 系数的所有系数。

79	65	60	70	77	68	58	75
85	71	64	59	55	61	65	83
87	79	69	68	65	76	78	94

原图像包含 256 或 2^8 个可能的灰度级，因此，编码过程从对原子图像的像素层次移动 -2^7 或 -128 个灰度级开始。得到的移位阵列为：

-76	-73	-67	-62	-58	-67	-64	-55
-65	-69	-62	-38	-19	-43	-59	-56
-66	-69	-60	-15	16	-24	-62	-55
-65	-70	-57	-6	26	-22	-58	-59
-61	-67	-60	-24	-2	-40	-60	-58
-49	-63	-68	-58	-51	-65	-70	-53
-43	-57	-64	-69	-73	-67	-63	-45
-41	-49	-59	-60	-63	-52	-50	-34

对 $N=8$，根据式(8.5.24)和式(8.5.32)的正向 DCT，变换的阵列为：

-415	-29	-62	25	55	-20	-1	3
7	-21	-62	9	11	-7	-6	6
-46	8	77	-25	-30	10	7	-5
-50	13	35	-15	-9	6	0	3
11	-8	-13	-2	-1	1	-4	1
-10	1	3	-3	-1	0	2	-1
-4	-1	2	-1	2	-3	1	-2
-1	-1	-1	-2	-1	-1	0	-1

表 8.17 JPEG 系数编码分类

范围	DC 差异类	AC 类
0	0	N/A
-1,1	1	1
-3,-2,2,3	2	2
-7,…,-4,4,…,7	3	3
-15,…,-8,8,…,15	4	4

（续表）

范　围	DC 差异类	AC 类
-31,…,-16,16,…,31	5	5
-63,…,-32,32,…,63	6	6
-127,…,-64,64,…,127	7	7
-255,…,-128,128,…,255	8	8
-511,…,-256,256,…,511	9	9
-1 023,…,-512,512,…,1 023	A	A
-2 047,…,-1 024,1 024,…,2 047	B	B
-4 095,…,-2 048,2 048,…,4 095	C	C
-8 191,…,-4 096,4 096,…,8 191	D	D
-16 383,…,-8 192,8 192,…,16 383	E	E
-32 767,…,-16 384,16 384,…,32 767	F	N/A

表 8.18　JPEG 默认 DC 编码（亮度）

分　类	基础编码	长　度	分　类	基础编码	长　度
0	010	3	6	1110	10
1	011	4	7	11110	12
2	100	5	8	111110	14
3	00	5	9	1111110	16
4	101	7	A	11111110	18
5	110	8	B	111111110	20

表 8.19　JPEG 默认 AC 编码（亮度）

扫描段/类	基础编码	长　度	扫描段/类	基础编码	长　度
0/0	**1010（= EOB）**	**4**			
0/1	00	3	8/1	11111010	9
0/2	01	4	8/2	111111111000000	17
0/3	100	6	8/3	1111111110110111	19
0/4	1011	8	8/4	1111111110111000	20
0/5	11010	10	8/5	1111111110111001	21
0/6	111000	12	8/6	1111111110111010	22
0/7	1111000	14	8/7	1111111110111011	23
0/8	1111110110	18	8/8	1111111110111100	24
0/9	1111111110000010	25	8/9	1111111110111101	25
0/A	1111111110000011	26	8/A	1111111110111110	26
1/1	1100	5	9/1	111111000	10
1/2	111001	8	9/2	1111111110111111	18
1/3	1111001	10	9/3	1111111111000000	19
1/4	111110110	13	9/4	1111111111000001	20
1/5	11111110110	16	9/5	1111111111000010	21
1/6	1111111110000100	22	9/6	1111111111000011	22

(续表)

扫描段/类	基础编码	长度	扫描段/类	基础编码	长度
1/7	1111111110000101	23	9/7	1111111111000100	23
1/8	1111111110000110	24	9/8	1111111111000101	24
1/9	1111111110000111	25	9/9	1111111111000110	25
1/A	1111111110001000	26	9/A	1111111111000111	26
2/1	11011	6	A/1	111111001	10
2/2	11111000	10	A/2	1111111111001000	18
2/3	1111110111	13	A/3	1111111111001001	19
2/4	1111111110001001	20	A/4	1111111111001010	20
2/5	1111111110001010	21	A/5	1111111111001011	21
2/6	1111111110001011	22	A/6	1111111111001100	22
2/7	1111111110001100	23	A/7	1111111111001101	23
2/8	1111111110001101	24	A/8	1111111111001110	24
2/9	1111111110001110	25	A/9	1111111111001111	25
2/A	1111111110001111	26	A/A	1111111111010000	26
3/1	111010	7	B/1	111111010	10
3/2	111110111	11	B/2	1111111111010001	18
3/3	11111110111	14	B/3	1111111111010010	19
3/4	1111111110010000	20	B/4	1111111111010011	20
3/5	1111111110010001	21	B/5	1111111111010100	21
3/6	1111111110010010	22	B/6	1111111111010101	22
3/7	1111111110010011	23	B/7	1111111111010110	23
3/8	1111111110010100	24	B/8	1111111111010111	24
3/9	1111111110010101	25	B/9	1111111111011000	25
3/A	1111111110010110	26	B/A	1111111111011001	26
4/1	111011	7	C/1	1111111010	11
4/2	1111111000	12	C/2	1111111111011010	18
4/3	1111111110010111	19	C/3	1111111111011011	19
4/4	1111111110011000	20	C/4	1111111111011100	20
4/5	1111111110011001	21	C/5	1111111111011101	21
4/6	1111111110011010	22	C/6	1111111111011110	22
4/7	1111111110011011	23	C/7	1111111111011111	23
4/8	1111111110011100	24	C/8	1111111111100000	24
4/9	1111111110011101	25	C/9	1111111111100001	25
4/A	1111111110011110	26	C/A	1111111111100010	26
5/1	1111010	8	D/1	11111111010	12
5/2	1111111001	12	D/2	1111111111100011	18

(续表)

扫描段/类	基础编码	长度	扫描段/类	基础编码	长度
5/3	1111111110011111	19	D/3	1111111111100100	19
5/4	1111111110100000	20	D/4	1111111111100101	20
5/5	1111111110100001	21	D/5	1111111111100110	21
5/6	1111111110100010	22	D/6	1111111111100111	22
5/7	1111111110100011	23	D/7	1111111111101000	23
5/8	1111111110100100	24	D/8	1111111111101001	24
5/9	1111111110100101	25	D/9	1111111111101010	25
5/A	1111111110100110	26	D/A	1111111111101011	26
6/1	1111011	8	E/1	111111110110	13
6/2	11111111000	13	E/2	1111111111101100	18
6/3	1111111110100111	19	E/3	1111111111101101	19
6/4	1111111110101000	20	E/4	1111111111101110	20
6/5	1111111110101001	21	F/5	1111111111101111	21
6/6	1111111110101010	22	E/6	1111111111110000	22
6/7	1111111110101011	23	F/7	1111111111110001	23
6/8	1111111110101100	24	E/8	1111111111110010	24
6/9	1111111110101101	25	E/9	1111111111110011	25
6/A	1111111110101110	26	E/A	1111111111110100	26
7/1	11111001	9	**F/0**	**111111110111**	**12**
7/2	11111111001	13	F/1	1111111111110101	17
7/3	1111111110101111	19	F/2	1111111111110110	18
7/4	1111111110110000	20	F/3	1111111111110111	19
7/5	1111111110110001	21	F/4	1111111111111000	20
7/6	1111111110110010	22	F/5	1111111111111001	21
7/7	1111111110110011	23	F/6	1111111111111010	22
7/8	1111111110110100	24	F/7	1111111111111011	23
7/9	1111111110110101	25	F/8	1111111111111100	24
7/A	1111111110110110	26	F/9	1111111111111101	25
			F/A	1111111111111110	26

如果用图 8.37(b)JPEG 推荐的标准化阵列对变换阵列进行量化,则进行按比例舍入[即,根据式(8.5.40)进行标准化]后的系数是:

$$-26 \quad -3 \quad -6 \quad 2 \quad 2 \quad 0 \quad 0 \quad 0$$

1	−2	−4	0	0	0	0	0
−3	1	5	−1	−1	0	0	0
−4	1	2	−1	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

这里,DC 系数使用下式计算:

$$\begin{aligned}\hat{T}(0,0) &= \mathrm{round}\left[\frac{T(0,0)}{Z(0,0)}\right]\\ &= \mathrm{round}\left[\frac{-415}{16}\right] = -26\end{aligned}$$

注意,变换和标准化处理过程产生了大量的零值系数。当根据图 8.36(d)的 z 形模式对这些系数进行重新排列的时候,得到的一维系数序列是:

[−26 −3 1 −3 −2 −6 2 −4 1 −4 1 1 5 0 2 0 0 −1 2 0 0 0 0 0 −1 −1 EOB]

这里,EOB 字符代表块结尾标志。使用的专用 EOB 霍夫曼码字(见表 8.19 中的类 0 和行程长度 0)表明在重新排列的序列中系数余项均为 0。

对于重排列系数序列默认的 JPEG 编码的构造,是从计算当前的 DC 系数和前面经过编码的子图像的 DC 系数之间的差开始的。由于这里使用的子图像来自图 8.23,并且紧挨着此子图像左边的经过变换和量化的子图像的 DC 系数值为 −17,所以得到的 DPCM 差为 [−26−(−17)]或 −9,此值位于表 8.17DC 差异类 4 中。根据表 8.18 的默认霍夫曼差异码,对一个类 4 的差的合理基础编码是 101(一个 3 比特编码),而经过完整编码的类 4 系数的总长度应有 7 比特。余下的 4 比特应该根据差值的最低有效位(LSB)生成。对一个一般的 DC 差异类(如,类 K),需要额外的 K 比特,并将正差或负差的 K 个 LSB 进行减 1 运算。对差为 −9 的值,合理的 LSB 为(0111)−1 或 0110,且完整的 DPCM 编码的 DC 码字是 1010110。

重排阵列的非零 AC 系数根据表 8.17 和表 8.19 可以进行类似的编码。主要差别在于每个默认的 AC 霍夫曼码字依赖于前述对非零系数编码后的零值系数的数目,以及非零系数的量级分类(见表 8.19 中的标记为扫描段/类的列)。因此,第一个重排阵列的非零 AC 系数(−3)的编码为 0100。这个编码的前两位表明这个系数属于量级分类 2 且没有零值系数在它的前面(见表 8.17);最后两位的生成使用了与获得 DC 差异编码的 LSB 所使用的相同处理过程。继续使用这种方法,完整的编码(重排列)阵列为:

1010110 0100 001 0100 0101 100001 0110 100011 001 100011 001
001 100101 11100110 110110 0110 11110100 000 1010

这里分别插入了空格以便于阅读。尽管在这个例子中并不需要这个阵列,但默认的 JPEG

编码包含对一个0后面跟着15个0的段设立的专用码字(见表8.19中的类0和行程长度F)。在完全编过码的重排阵列中的总比特数(在这个例子中,表示整幅8×8大小的8位子图像所要求的比特数)是92。得到的压缩率为512/92,或大约5.6:1。

为了解压缩一幅压缩后的JPEG子图像,解码器必须首先重新产生标准化的导出压缩比特流的变换系数。因为一个霍夫曼编码的二值序列具有瞬时性并且其解码具有惟一性,所以用简单的查询表方式就可以很容易地实现这一步骤。这里量化系数的重新生成的阵列为:

−26	−3	−6	2	2	0	0	0
1	−2	−4	0	0	0	0	0
−3	1	5	−1	−1	0	0	0
−4	1	2	−1	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

在根据式(8.5.42)进行反标准化之后,阵列变成:

−416	−33	−60	32	48	0	0	0
12	−24	−56	0	0	0	0	0
−42	13	80	−24	−40	0	0	0
−56	17	44	−29	0	0	0	0
18	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

这里,例如DC系数,可以根据下式计算:

$$\dot{T}(0,0)=\hat{T}(0,0)Z(0,0)=(-26)(16)=-416$$

完整的重构子图像通过式(8.5.25)和式(8.5.32)得到的反向标准化阵列的逆DCT得出:

−70	−64	−61	−64	−69	−66	−58	−50
−72	−73	−61	−39	−30	−40	−54	−59

-68	-78	-58	-9	13	-12	-48	-64
-59	-77	-57	0	22	-13	-51	-60
-54	-75	-64	-23	-13	-44	-63	-56
-52	-71	-72	-54	-54	-71	-71	-54
-45	-59	-70	-68	-67	-67	-61	-50
-35	-47	-61	-66	-60	-48	-44	-44

而对每个逆变换像素用 $+2^7$(或 +128)进行移位得到:

58	64	67	64	59	62	70	78
56	55	67	89	98	88	74	69
60	50	70	119	141	116	80	64
69	51	71	128	149	115	77	68
74	53	64	105	115	84	65	72
76	57	56	74	75	57	57	74
83	69	59	60	61	61	67	78
93	81	67	62	69	80	84	84

原子图像和重构子图像之间的任何差异都是 JPEG 压缩和解压缩处理过程固有的有损性造成的。在这个例子中,误差范围从 -14 到 +11,并且呈如下的分布状态:

-6	-9	-6	2	11	-1	-6	-5
7	4	-1	1	11	-3	-5	3
2	9	-2	-6	-3	-12	-14	9
-6	7	0	-4	-5	-9	-7	1
-7	8	4	-1	11	4	3	-2
3	8	4	-4	2	11	1	1
2	2	5	-1	-6	0	-2	5
-6	-2	2	6	-4	-4	-6	10

整体压缩和重构处理的均方根误差接近 5.9 个灰度级。

前面例子中的重构子图像大约在以图 8.38(a)中妇女的右眼为中心的位置上。这里请注意,在原始子图像和重构子图像上,第 4 行第 5 列的位置上都含有一个局部灰度级的波峰,这

个位置是妇女的瞳孔反射的一个亮点。这个局部波峰造成重构子图像的均方根误差明显超过了完全解压缩图像的总体误差。实际上,这个误差接近图 8.38(a)中误差的两倍。图 8.38(a)的图像也是使用基准 JPEG 算法压缩得到的。原因是,原图像中的许多子图像都接近恒定并且描述这些图像时几乎没有失真。图 8.38(b)给出了另外一个使用 JPEG 基准压缩得到的结果。

JPEG 2000

尽管这种标准仍未被正式采用,但 JPEG 2000 相对于原始的 JPEG 标准,在对连续色调静止图像的压缩方面以及对压缩数据的访问方面提供了更大的灵活性。例如,一幅按 JPEG 2000 标准压缩的图像的一些部分可以从图像中提取出来进行转发、存储、显示和/或编辑。这个标准是以 8.5.3 节的小波编码技术为基础的。系数量化与单一尺度和子带相适应,并且量化的系数以位平面为基础(见 8.4 节)进行算术编码。使用标准符号,一幅图像可以进行如下(ISO/IEC[2000])编码。

编码处理的第一步是通过减 2^{Ssiz-1} 将被编码的 *Ssiz* 比特无符号图像样值向 DC 级移动。如果图像具有不止一个组成分量——比如彩色图像的红色、绿色和蓝色面——则对每个分量进行单独的移动。如果有 3 个精确的分量,则可以使用可逆的或非可逆的线性组合进行任意选择的解相关处理。标准中不可逆的分量变换是:

$$\begin{aligned} Y_0(x,y) &= 0.299I_0(x,y)+0.587I_1(x,y)+0.114I_2(x,y) \\ Y_1(x,y) &= -0.16875I_0(x,y)-0.33126I_1(x,y)+0.5I_2(x,y) \\ Y_2(x,y) &= 0.5I_0(x,y)-0.41869I_1(x,y)-0.08131I_2(x,y) \end{aligned} \qquad (8.6.1)$$

这里 I_0,I_1 和 I_2 是级别的平移输入分量,Y_0,Y_1 和 Y_2 是相应的解相关分量。如果输入分量是彩色图像的红色、绿色和蓝色平面,式(8.6.1)用 $Y'C_bC_r$ 彩色视频变换来近似 $R'G'B'$(Poynton[1996])①。变换的目的是改善压缩效率;变换后的分量 Y_1 和 Y_2 是差值图像,此图像的直方图在零点附近具有很高的峰。

在图像进行了水平移动和选择性的解相关处理之后,它的分量被分割成像块。这些块是像素的矩形阵列,而这些像素包含着所有分量相同的相关部分。因此,像块处理生成了可以独立进行提取和重构的块分量,这种处理提供了一种对编码图像的某一有限区域访问和/或操作的简单原理。

对每个块分量的行和列的一维离散小波变换就可以计算了。对于无误差压缩,这种变换是以双正交、5-3 系数尺度和小波向量(Le Gall 和 Tabatabai[1988])为基础的。对于非整数值变换系数还定义了一个四舍五入的程序。在有损应用中,使用 9-7 小波向量尺度系数(Antonini,Barlaud,Mathieu 和 Daubechies[1992])。无论哪种情况,都是使用 7.4 节的快速小波变换或通过一种基于提升的方法进行计算的。表 8.20 中给出了构造一个 9-7 FWT 分析滤波器组所需的系数。补充的基于提升的实现涉及 6 种连续的"提升"和"尺度"操作:

---

① $R'G'B'$ 是线性 CIE(International Commission on Illumination,国际照明委员会)*RGB* 色度值经图像矫正且非线性的版本。Y' 是亮度,C_b 和 C_r 是色差(即,标定的 $B'-Y'$ 和 $R'-Y'$ 值)。

$$
\begin{aligned}
&Y(2n+1)=X(2n+1)+\alpha[X(2n)+X(2n+2)], i_0-3\leqslant 2n+1<i_1+3\\
&Y(2n)=X(2n)+\beta[Y(2n-1)+Y(2n+1)],\ i_0-2\leqslant 2n<i_1+2\\
&Y(2n+1)=Y(2n+1)+\gamma[Y(2n)+Y(2n+2)], i_0-1\leqslant 2n+1<i_1+1\\
&Y(2n)=Y(2n)+\delta[Y(2n-1)+Y(2n+1)],\ i_0\leqslant 2n<i_1\\
&Y(2n+1)=-K\cdot Y(2n+1), i_0\leqslant 2n+1<i_1\\
&Y(2n)=Y(2n)/K, i_0\leqslant 2n<i_1
\end{aligned}
\tag{8.6.2}
$$

这里,X 是被变换的块分量,Y 是变换结果,i_0 和 i_1 定义了分量内的块分量的位置。即,它们是将要变换的块分量行和列的第一个取样及紧接着的后一个取样的索引。变量 n 是以 i_0,i_1 为基础计算的,并且将形成6种操作。如果 $n<i_0$ 或 $n\geqslant i_1$,则 $X(n)$ 通过对称地对 X 进行扩展而得到。例如,$X(i_0-1)=X(i_0+1)$,$X(i_0-2)=X(i_0+2)$,$X(i_1)=X(i_1-2)$,$X(i_1+1)=X(i_1-3)$。在关于提升和尺度操作的总结部分,Y 的偶下标的值与 FWT 低通滤波输出相等;Y 的奇数下标的值与高通 FWT 滤波输出相对应。提升参量 α,β,γ 和 δ 分别为 −1.586 134 342,−0.052 980 118,0.882 911 075 和 0.433 506 852。尺度因子 K 为 1.230 174 105。

表 8.20 不可逆的 9-7 小波变换的低通和高通分析滤波器的脉冲响应

滤波器抽头	高通小波系数	低通尺度系数
0	−1.115 087 052 456 994	0.602 949 018 236 357 9
±1	0.591 271 763 114 247 0	0.266 864 118 442 872 3
±2	0.057 543 526 228 499 57	−0.078 223 266 528 987 85
±3	−0.091 271 763 114 249 48	−0.016 864 118 442 874 95
±4	0	0.026 748 757 410 809 76

刚才描述的变换生成了4个子带——分别是块分量的低分辨率近似和分量的水平、垂直及对角频率特征。后续迭代限制前面分解的近似值系数,将这种变换重复 N_L 次可以得到一个 N_L 尺度的小波变换。相邻尺度的变换联系到2的幂次空间且最低尺度仅包含显式定义的原始块分量的近似。正如在图8.46中推测的那样,这里总结了 $N_L=2$ 时的标准符号,一个一般的 N_L 尺度变换包含了 $3N_L+1$ 个子带,而这些子带系数,对于 $b=N_LLL, N_LHL, \cdots, 1HL, 1LH, 1HH$,用 a_b 代表。这个标准没有指明需要计算的尺度的数目。

当每个块分量都经过处理之后,变换系数的总数等于原始图像中的取样数目,但重要的可视信息被集中于少数系数中。为了减少表示变换的数据比特数,子带 b 的系数 $a_b(u,v)$ 使用下式量化为值 $q_b(u,v)$:

$$
q_b(u,v)=\operatorname{sign}[a_b(u,v)]\cdot\operatorname{floor}\left[\frac{|a_b(u,v)|}{\Delta_b}\right] \tag{8.6.3}
$$

这里量化步长 Δ_b 为:

$$
\Delta_b=2^{R_b-\varepsilon_b}\left(1+\frac{\mu_b}{2^{11}}\right) \tag{8.6.4}
$$

R_b 是子带 b 的标定动态范围,而 ε_b 和 μ_b 是分配给子带系数的指数和尾数的比特数。子带 b 的

标定动态范围是,用于表示原始图像的比特数和对子带 b 的分析增益比特的位数之和。子带分析增益比特遵从图 8.46 中显示的简单模式。例如,对子带 $b=1HH$ 存在两个分析增益比特。

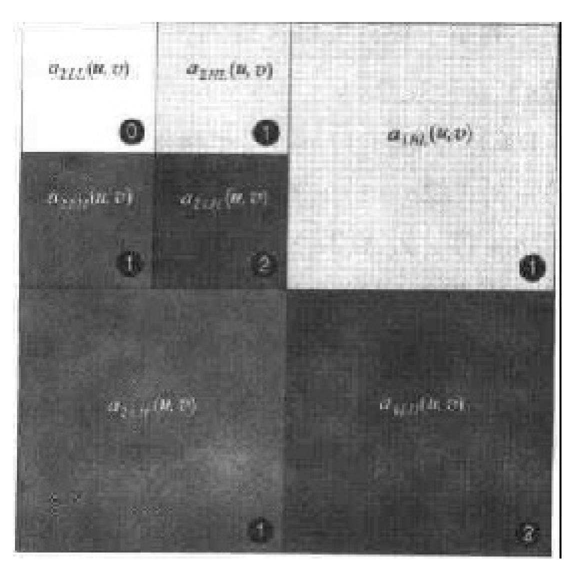


图 8.46 JPEG 2000 的 2 尺度小波变换块分量系数符号和分析结果

对无误差压缩,$\mu_b=0, R_b=\varepsilon_b, \Delta_b=1$。对不可逆压缩,在标准中没有指定特定的量化步长。相反,必须以子带为基础向解码器提供指数和尾数的比特数,这被称为显式量化,或仅对 N_LLL 子带,称为隐式量化。对于后一种情况,对余下的子带使用推测的 N_LLL 子带参数进行量化。令 ε_0 和 μ_0 表示分配到 N_LLL 子带的位数,对子带 b 推测的参数为:

$$\begin{aligned}\mu_b &= \mu_0 \\ \varepsilon_b &= \varepsilon_0 + nsd_b - nsd_0\end{aligned} \tag{8.6.5}$$

这里,nsd_b 表示由原始图像块分量到子带 b 的子带分解层次的数目。

编码过程的最后一步是系数比特建模、算术编码、比特流分层和分组。每个变换块分量的子带系数排列在称为码块的矩形块中,这个块一次对一个位平面进行独立编码。从带有非零元素的最高有效位平面开始,每个位平面进行 3 次处理。位平面的每比特仅在 3 次处理中的一次进行编码。其过程分别称为有效传播、量级细化和净化。然后,对输出进行算术编码并从其他编码块使用相似的途径进行分组以形成层。一个层是来自每个编码块的编码途径分组的任意数。得到的层被最终分割成信息包,提供从总体编码流中提取某一空间区域的附加方法。包是编码码流的基本单元。

JPEG 2000 解码器是前述过程的简单反操作。对比特建模、算术编码、分层和编码流打包进行解码之后,对原始图像块分量的用户选择的数重建子带。尽管编码器可能已经在一个特定的子带上对 M_b 位平面进行了编码,但用户可以选择只用 N_b 位平面进行解码,这取决于嵌入的编码流的性质。这实际上是使用 $2^{M_b-N_b}\cdot\Delta b$ 的步长对编码块系数进行量化。所有未解码的比特都设置为零,并且对得到的表示为 $\overline{q}_b(u,v)$ 的系数使用下式进行反量化:

$$R_{q_b}(u,v)=\begin{cases}(\overline{q}_b(u,v)+2^{M_b-N_b(u,v)})\cdot\Delta_b & \overline{q}_b(u,v)>0\\(\overline{q}_b(u,v)-2^{M_b-N_b(u,v)})\cdot\Delta_b & \overline{q}_b(u,v)<0\\0 & \overline{q}_b(u,v)=0\end{cases} \tag{8.6.6}$$

这里 $R_{q_b}(u,v)$表示一个反向量化变换系数,而 $N_b(u,v)$是针对 $\overline{q}_b(u,v)$的解码位平面的数目。之后,反向量化系数使用一个 FWT^{-1}滤波器组按列和按行进行逆向变换,这些系数根据表 8.20和式(7.1.15)或通过下列基于提升的操作得到:

$$
\begin{aligned}
&X(2n) = K\cdot Y(2n),\ i_0-3\leqslant 2n < i_1+3\\
&X(2n+1) = (-1/K)\cdot Y(2n+1),\ i_0-2\leqslant 2n+1 < i_1+2\\
&X(2n) = X(2n) - \delta[X(2n-1)+X(2n+1)],\ i_0-3\leqslant 2n < i_1+3\\
&X(2n+1) = X(2n+1) - \gamma[X(2n)+X(2n+2)],\ i_0-2\leqslant 2n+1 < i_1+2\\
&X(2n) = X(2n) - \beta[X(2n-1)+X(2n+1)],\ i_0-1\leqslant 2n < i_1+1\\
&X(2n+1) = X(2n+1) - \alpha[X(2n)+X(2n+2)],\ i_0\leqslant 2n+1 < i_1
\end{aligned}
\tag{8.6.7}
$$

这里,参数 $\alpha,\beta,\gamma,\delta$ 和 K 同式(8.6.2)中的定义一样。经过反向量化的系数行或列元素 $Y(n)$ 在需要的时候可以进行对称的扩充。解码的最后一步是分量块的组合、反向分量变换(如果需要的话)和 DC 水平移动。对于不可逆编码,反向分量变换是:

$$
\begin{aligned}
&I_0(x,y) = Y_0(x,y) + 1.402Y_2(x,y)\\
&I_1(x,y) = Y_0(x,y) - 0.34413Y_1(x,y) - 0.71414Y_2(x,y)\\
&I_2(x,y) = Y_0(x,y) + 1.772Y_1(x,y)
\end{aligned}
\tag{8.6.8}
$$

且将变换后的像素用 $+2^{S_{siz}-1}$ 移动。8.5.3 节的图 8.40 和图 8.41 显示的压缩率为 34:1 到 167:1。这些图像是使用不可逆的 JPEG 2000 算法生成的。

8.6.3　视频压缩标准

视频压缩标准将前面章节的以变换为基础的静止图像压缩技术扩展成包括某些减少暂时性和图像帧间冗余的方法。尽管现今使用的视频编码标准有好几种,但其中大多数依赖相似的视频压缩技术。根据所面对的应用领域的不同,可以将这些标准分为两个大类:(1)电视会议标准和(2)多媒体标准。

许多电视会议标准[包括 H.261(也称为 PX64),H.262,H.263 和 H.320]已经由国际电信联盟(ITU)制定了。国际电信联盟(ITU)是 CCITT 的后继组织。H.261 计划以可承担的远程通信比特率操作,并支持在 T1① 线路上以小于 150 ms 的延迟进行全动态视频传输。超过 150 ms 的延迟时间就不能给观察者以直接视觉反馈的感觉。另一方面,H.263 是设计用来为极低比特率视频服务的,它应用的比特率在 10~30 Kbps 之间,一种 H.261 标准的扩充是为综合业务数字网② (ISDN)的带宽而设的。每种标准都采用一套运动补偿,基于 DCT 的编码方案。由于运动估计在变换领域难于实现,所以将这些像素块(称为宏块)与前面图像帧的相邻块进行比较,并用于计算一个运动补偿预测误差。然后将预测误差在 8×8 的像素块内进行离散余弦变换,并为了传输或存储进行量化和编码。

针对视频点播、数字 HDTV 广播和图像/视频数据库服务等领域的多媒体视频压缩标准使

---

① T1 线路是由贝尔系统为了在 10~50 mile 的短距离内进行数字语音通信而提出的。它具有 24 路电话分时多路复用,脉冲调制,并按照在单一 T1 线路上传输的 1.544 Mbps 的 PCM(脉冲编码解调)信号进行编码(1 mile = 1 609.344 m)。

② 两个 ISDN"B"通道提供足够的带宽(即,128 Kbps)以每秒 15 帧的速度传输压缩的 320×240 大小的图像。

用相似的运动估计和编码技术。其中主要的标准——MPEG-1,MPEG-2和MPEG-4——在 CCITT 和 ISO 的动态图像专家组的支持下得以建立。MPEG-1 是一种用于存储和检索类似光盘只读存储器(CD-ROM)的数字媒体视频的“娱乐品质”的编码标准,它支持的比特率接近1.5 Mbps。MPEG-2 标准用于涉及 NTSC/PAL[①] 和 CCIR601[②] 之间的视频品质的应用。它支持比特率为 2~10 Mbps的传送,这一范围很适合于有线电视传送和窄信道卫星广播。制定 MPEG-1 和 MPEG-2 的目的是高效率地存储和传输数字音频和视频(AV)资料。另一方面,MPEG-4 标准提供了:(1)改善的视频压缩效率;(2)基于内容的交互性,比如面向对象的 AV 访问和高效率的自然数据与人工数据的综合;(3)全球访问,包括在易于出现差错的环境下改善传输的稳健性,增加或撤销 AV 对象,以及变换对象分辨率。尽管这些功能产生了任意分割有形视频对象的需求,但分割不是标准的一部分。大量的视频内容——像计算机游戏——是以视频对象的形式产生和存在的。MPEG-4 以移动通信和公共交换电话网(PSTN)应用的 5 Kbps 和64 Kbps之间的比特率以及 TV 和电影应用的 4 Mbps 的比特率为目标。另外,这种标准还支持固定比特率和可变比特率编码。

类似 ITU 电视会议标准,MPEG 标准是围绕一种混合的以块为基础的 DPCM/DCT 编码方案建立的。图 8.47 显示了一种典型的 MPEG 编码器。这个编码器利用了视频帧内部的和相邻视频帧之间的冗余、视频帧之间的运动一致性以及人类视觉系统的心理视觉特性。编码器的输入是一个 8×8 的像素阵列,这个阵列称为一个图像块。标准规定一个宏块为一个 2×2 的图像块阵列(即,一个 16×16 的图像元素阵列),一个片段是一行不重叠的宏块。对于彩色视频图像,一个宏块由 4 个亮度块(它们分别表示为 Y_1 到 Y_4)及两个色度块 C_b 和 C_r 组成。回想一下,色差信号 C_b 是蓝色减去亮度,而 C_r 是红色减去亮度。因为眼睛对于颜色的空间判断能力远小于对于亮度的判断,所以这两个分量的取样通常只是亮度信号在水平和垂直方向分辨率的一半。在 $Y':C_b:C_r$ 之间的取样比率为 4:1:1。

对比图 8.47 中原来输入输出路径的灰度元素在 JPEG 编码器中的变换、量化和变长编码操作。主要的差别在于输入,输入可以是一个常规的图像数据块或者是常规的图像块和以前面的和/或后续视频帧中相似的图像数据块为基础所得的预测块之间的差。这种输入导致 3 种基本类型的编码输出帧:

1. 帧内或独立帧(I-帧)。I-帧对于它的前帧和它的后续视频帧是独立压缩的。就这 3 种可能出现的编码输出帧来说,I-帧与 JPEG 编码的图像具有最高的相似性。再有,它是生成后续的 P-帧和 B-帧所需的运动估计的参考点。I-帧提供了最高级的随机访问功能、编辑的简易性和最好的阻止传输误差扩大的能力。结果是在所有的标准中都需要在压缩的编码流中周期性地插入这些帧。

2. 预测帧(P-帧)。P-帧是经过压缩的当前帧和基于前一 I-帧或 P-帧的预测之间的差。这一差是在图 8.47 最左边的加法器中形成的。预测是运动补偿,一般涉及图 8.47 底部围绕着紧挨着的相邻块在当前帧中滑动的解码块,以及计算相关性的量度(如相邻像素间差的平方和)。实际上,这种处理通常按亚像素尺度的增量进行(比如一次将子图像移动 1/4 个像素),

① NTSC 和 PAL 分别是美国国家电视广播系统委员会(National Television System Committee)和逐行倒相(Phase Alternate Line)的首字母缩写。它们都是合成彩色视频标准。

② CCIR 是国际无线电顾问委员会的首字母缩写。

这使得必须在计算相关性量度之前先计算内插像素值。计算得出的运动向量被变长编码,并且作为编码数据流整体的一部分进行传送。运动估计在宏块水平上进行。

3. 双向帧(B-帧)。B-帧是对当前帧和基于前一个 I-帧或 P-帧以及下一个 P-帧的预测之间的差进行编码。因此,解码器必须访问过去的和将来的参考帧。所以编码帧在传送之前要进行重排;而解码器在适当的序列中将这些帧进行重构和显示。

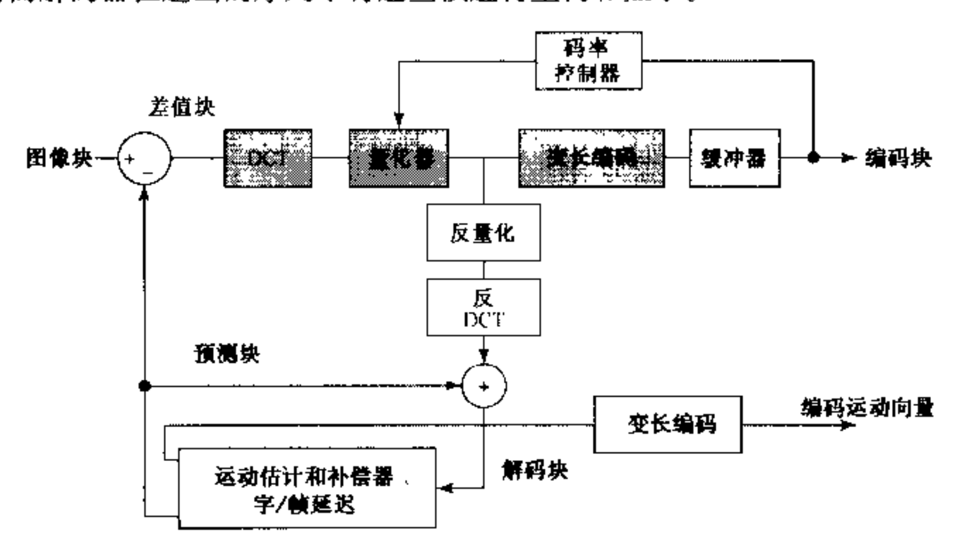


图 8.47 一种基本的运动补偿视频压缩的 DPCM/DCT 编码器

图 8.47 的编码器设计用来生成一个比特流,这个比特流与预定的视频信道的容量相匹配。为了实现这样的匹配,必须将量化参数作为输出缓冲区的占用率的函数通过速率控制器进行调整。当缓冲区越来越满的时候,量化就比较粗略,这样就没有多少比特流送入缓冲区。

小结

这一章的主要目的是介绍数字图像压缩的理论基础,并描述最常用的压缩方法,这些压缩方法在当前构成了已有的这类技术的核心。尽管我们的论述实际上是介绍性的,但所使用的材料的深度和广度足以适合作为这一主题领域阅读的基础。所列的参考资料可以作为一种入门材料,继而可以广泛涉猎论述图像压缩和相关主题的著述。除了涉及灰度级成像等方面的应用之外,压缩方法在文献图像存储和传送方面也日益扮演着越来越重要的角色,8.6 节中讨论的各种国际标准的出现就是很好的例证。除了医学成像,压缩也是少数几个根据商业上的广泛要求对普遍接受的标准进行授权使用的图像处理领域之一。

参考资料

本章中的(一般限制在 8.1 节和 8.2 节中)介绍性材料作为图像压缩的基础在大多数一般的图像处理方面的著述中都可以以不同的形式见到,有关的这些著述在第 1 章的结尾处都有引证。8.1.3 节中关于改善灰度级量化方面的材料是以 Bisignani, Richards 和 Whelan[1996]的著述为基础的。关于人类视觉系统的附加信息见 Netravali 和 Limb[1980]的著述,此外还有

Huang[1966],Schreiber 和 Knapp[1958]的著述和第 2 章结尾处引证的参考资料。主观保真度准则在 Frendendall 和 Behrend[1960]中做了讨论。关于交换或有限自动机理论的介绍性的文章,以及一般的信息论文章中都涵盖了误差检测和修正编码方面的内容。

8.3 节中的材料是以几本关于信息论的非常优秀的著作为基础的。值得注意的是 Abramson[1963],Blahut[1987]和 Berger[1971]的相关著述。香农的经典文章"通信中的数学理论"[1948]是本小节中大多数材料的基础,也是另一份优秀的参考资料。

8.4 节中的无误差编码技术的论述很大程度上是以叙述中引证的原始文章为基础的。小节中所包含的算法是这一领域中具有代表性的成果,但决不是全部的成果。关于 LZW 编码的材料来自 Ziv 和 Lempel[1977,1978]所做的原始工作。关于算术编码的材料是 Witten,Neal 和 Cleary[1987]的著述的进一步发展。在 Pennebaker 等人[1988]的著述中对一种算术编码的更重要的实现方法进行了概述。关于位平面编码的其他一些信息可以参见 Schwartz 和 Barker[1966]的著述与 Rabbani 和 Jones[1991]所著的指南。在这本指南中同样包含了一段很好的关于无损预测编码的论述。Huang 和 Hussian[1975] 首先发表了有关白块跳行的细节。相对地址编码和预测差值量化最初分别由 Yamazaki,Wakahara 和 Teramura[1976]以及 Huang 和 Tretiak[1972]发表。式(8.5.19)的自适应预测器来自 Graham[1958]的著述。

8.5 节中的材料涵盖了主要的有损编码技术。许多种方法都是直接以这些技术为基础的。在这些技术中值得注意的是混合编码(Habibi[1974]),这种方案组合了一维编码和 DPCM,得到了与计算量较少的二维变换编码相同的效果;子带编码(Woods 和 O'Neil[1986])将一幅图像过滤为一组图像(用不同的空间频率)。这组图像可以进行独立的 DPCM 编码;还有帧间编码(Roese 等[1977]),使用这种方法时,在一个图像的时间序列中通过使用一个预测或变换编码方法减少连续帧之间的冗余。另外,许多有损技术与我们讲述的技术有很紧密的联系。其中的一些技术包括块截取编码(Delp 和 Mitchell[1979]),在这种技术中,对每个 $n \times n$ 大小的细分图像的块设计一个 1 比特量化器;向量量化(Linde 等[1980])技术将一幅图像分解为向量(包含像素、变换系数,等等),这些向量按一个可能出现的向量的码书进行匹配,并用编码表示最佳拟合;分层编码(Knowlton[1980])通常涉及构造一种金字塔结构的数据集合,可以对这种数据集合进行渐进式的访问以便得到初始图像越来越好的表示。这些参考资料不需要提到这些技术的发明人;它们给出了关于这些方法的额外阅读的一个起点。其他重要的文章或书籍包括 Tasto 和 Wintz[1971],Gharavi 和 Tabatabai[1988],Baylon 和 Lim[1990],Candy 等人[1971],Jain 和 Jain[1981],Healy 和 Mitchell[1981],Lema 和 Mitchell[1984],Udpikar 和 Raina[1987],Gray[1984],Equitz[1989],Sezan 等人[1989],Tanimoto[1979],Blume 和 Fand[1989],Rabbani 和 Jones[1991],以及 Storer 和 Reif[1991]的著述。几乎每一期 IEEE 图像处理学报都包含几篇关于视频和静止图像压缩的文章,及许多与以小波和分形为基础的压缩、矢量量化、视频运动补偿相关的文章。例如,Boulgouris 等人[2001],Martin 和 Bell[2001],Chen 和 Wilson[2000],Hartenstein 等人[2000],Yang 和 Ramchandran[2000],以及 Meyer 等人[2000]的著述都可以作为进一步阅读和参考的起点。

8.6 节主要以国际标准组织与国际电话和电报咨询委员会发表的草案和正式标准为主要内容。这些来自标准组织或美国国家标准联盟(ANSI)的文献都是可以得到的。在压缩标准方面其他的参考文献包括 Hunter 和 Robinson[1980],Ang 等人[1991],Fox[1991],Pennebaker 和 Mitchell[1992],Bhatt 等人[1997],Sikora[1997],Bhaskaran 和 Konstantinos[1997],Ngan 等人

[1999],Weinberger 等人[2000],以及 Symes[2001]的相关著述。

还有几篇已经发表的有关图像压缩领域的综述文章。其中值得注意的是 Netravali 和 Limb [1980],A.K.Jain[1981]的文章,IEEE 通信学报[1981]关于图像通信系统的特刊,IEEE 会议论文集关于图形编码的特刊[1980],IEEE 会议论文集关于视觉通信系统的特刊[1985],IEEE 图像处理学报关于图像序列压缩的特刊[1994],以及 IEEE 图像处理学报关于向量量化的特刊[1996]。

习题

8.1 (a)变长编码程序能否用于压缩具有 2^n 个灰度级的直方图均衡过的图像?请解释。
(b)这样的图像是否包含可以用于数据压缩的像素间冗余?

8.2 8.1.2 节中描述了行程长度编码程序的一种变形。这种方法包括:(1)只对 0 或 1(不是 0 和 1 均出现)的行程进行编码,(2)在每条线的起点处分配一个指定的编码来减少传输误差的影响。一种可能出现的编码对是(x_k, r_k),这里 x_k 和 r_k 分别表示第 k 个行程的起始坐标和行程长度。编码(0,0)用于表示每条新的线。
(a)当对一幅 $2^n \times 2^n$ 大小的二值图像进行行程编码的时候,推导为确保数据压缩而要求的每条扫描线最大平均行程长度的一般表达式。
(b)计算 $n = 10$ 时的最大允许值。

★**8.3** 考虑灰度级数据{12,12,13,13,10,13,57,54}的一条 8 像素的线。这条线已经经过精度为 6 比特的均匀量化。构造它的 3 位 IGS 编码。

8.4 计算习题 8.3 中经过解码的 IGS 数据的 rms 误差和 rms 信噪比。

8.5 (a)使用 Hamming(7,4)编码对表 8.2 中的 IGS 量化数据进行编码。
★(b)如果存在的话,判断在 Hamming 编码的信息中哪一位是错误的:1100111,1100110 和 1100010。经过解码的值是什么?

★**8.6** 信息的以 e 为底的单元通常称为一个奈特(nat),底为 10 的信息单元称为一个哈特利(Hartley)。计算将这些单元与底为 2 的信息单元(比特)联系起来的转换因子。

★**8.7** 证明对于一个有 q 个符号的零记忆信源,熵的最大值为 $\log q$。当且仅当所有符号是等概率出现的时候,达到最大值。提示:考虑量 $\log q - H(\mathbf{z})$并注意不等式 $\ln x \leqslant x - 1$。

8.8 计算与 $A = \{0,1\}, B = \{0,1\}, \mathbf{z} = [0.75, 0.25]^T$ 和

$$\mathbf{Q} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{10} & \frac{9}{10} \end{bmatrix}$$

的信道相联系的各种概率。
包括 $P(a=0), P(a=1), P(b=0), P(b=1), P(b=0 \mid a=0)$,
$P(b=0 \mid a=1), P(b=1 \mid a=0), P(b=1 \mid a=1), P(a=0 \mid b=0)$,
$P(a=0 \mid b=1), P(a=1 \mid b=0), P(a=1 \mid b=1), P(a=0, b=0)$,
$P(a=0, b=1), P(a=1, b=0), P(a=1, b=1)$。

★**8.9** 考虑 8.3.2 节中的二值信息源和示例的 BSC,并令 $p_{bs} = 3/4, p_e = 1/3$。

(a)这个信源的熵是多少?

(b)当观察输出的时候,关于输入存在的不确定性是多大?

(c)这种不确定性的差别叫做什么? 并且如何与信道的容量进行数字上的比较?

8.10 一个二值删除信道是某一个被传输的符号没被接收到的概率为 β 的信道。这个信道有 3 种可能的输出:一个 0,一次删除(没接收到符号)和一个 1;这 3 个输出构成了二值删除信道矩阵的 3 行:

$$\mathbf{Q}=\begin{bmatrix}1-\beta & 0\\ \beta & \beta\\ 0 & 1-\beta\end{bmatrix}$$

(a)找到信道的容量。

★(b)希望得到一个出错概率为 0.125 的二值对称信道还是一个删除概率 $\beta=0.5$ 的删除信道?

8.11 一个零记忆高斯信源的率失真函数如下:

$$R(D)=\begin{cases}\dfrac{1}{2}\log\dfrac{\sigma^2}{D} & 0\leqslant D\leqslant\sigma^2\\ 0 & D\geqslant\sigma^2\end{cases}$$

★(a)绘制这个函数的曲线。

(b)$D_{\max}$是什么?

(c)如果要求失真不超过信源方差的 75%,则可以达到的最大压缩是多少?

8.12 (a)对一个三符号的信源有多少个惟一的霍夫曼编码?

(b)构造这些惟一的编码。

8.13 (a)计算表 8.1 中规定了符号概率的信源的熵。

(b)对信源符号构造霍夫曼编码,并解释构造的编码和表中的编码 2 之间的所有差异。

(c)为这个分布构造最佳的 B_1 编码。

(d)构造最佳的 2 位二值移位码。

(e)将符号分割为四个一块的两块,并构造最佳的霍夫曼移位码。

(f)对每个编码计算平均字长并将它们与(a)中的熵相比较。

★8.14 算术解码过程是编码过程的逆过程。对给出的编码模型信息 0.23355 进行解码。

符　号	概　率
a	0.2
e	0.3
i	0.1
o	0.2
u	0.1
!	0.1

8.15 使用 8.4.2 节中的 LZW 编码算法对 7 位 ASCII 字串"aaaaaaaaaaa"进行编码。

★8.16 为例 8.12 的 LZW 编码输出的解码操作设计一种算法。由于在编码期间使用的字典此时不能使用,所以在对输出进行解码的时候必须重建码书。

8.17 (a)构造完整的4位灰度编码。

(b)为将灰度编码的数字转换成与它等价的二进制值设计一个一般的程序,并使用它对0111010100111进行解码。

8.18 一个64×64像素大小的二值图像已经使用每个块4个像素的一维WBS进行了编码。图像一行的WBS编码是0110010000001000010010000000。这里0表示黑色像素。

(a)对该行进行解码。

(b)设计一个一维迭代WBS程序。这个程序从寻找所有的白色线条(一个64像素的块)开始并不断地对非白色的区间进行二等分,直到达到每个块4个像素。

(c)使用你的算法对前面的解码行进行编码。编码结果要求使用更少的位数。

8.19 ★(a)解释为什么在前面的直线上经过e的首次相似过渡在相对地址编码中用做c'。

(b)能设计一个替代的方法吗?

8.20 一幅图像的自相关函数是式(8.5.12)在$\rho_h = 0$时的形式。使用一个二阶预测因子对此图像进行DPCM编码。

(a)构造自相关矩阵**R**和向量**r**。

(b)找出最佳预测系数。

(c)计算使用最佳系数得到的预测误差的方差。

★8.21 推导出$L = 4$时的劳埃德·马克斯判定和重构层次,以及均匀概率密度函数:

$$p(s) = \begin{cases} \dfrac{1}{2A} & -A \leqslant s \leqslant A \\ 0 & \text{其他} \end{cases}$$

8.22 使用CCITT第4组压缩算法对下列双线线段的第2线段进行编码。

0 1 1 0 0 1 1 1 0 0 1 1 1 1 1 1 1 1 0 0 0 0 1

1 1 1 1 1 1 1 0 0 0 1 1 1 0 0 0 0 1 1 1 1 1 1

假设初始基准元素a_0被放置在第2条线段的第1个像素位置上。

★8.23 (a)列出JPEG DC系数差异类3的所有成员。

(b)使用表8.18计算默认的霍夫曼编码。

8.24 一位来自一所著名的研究性医院的放射学家最近参加了一次医学会议。在会上展示了一种可以通过标准T1电话线路传输$4\,096 \times 4\,096$大小的12位数字化X光图像的系统。这个系统以一种渐进的技术得到的一种压缩形式传送图像。这种技术在观察站上首先重构出X光的相当好的近似图像,然后经过逐步细化得到一幅无误差的显示图像。产生第一幅近似图像的数据传输要花5~6 s的时间。在接下来的1分钟之内,每隔5~6 s(平均)会进行一次细化。重构X光图像的过程中,第一次和最后一次细化分别对图像具有最大的和最小的影响。医生对这套系统的印象很好。因为可以使用首幅X光图像的近似图像开始进行诊断,而在生成了无误差X光重构图像的时候就可以完成诊断了。在回办公室的途中,她拟订了一份报送医院领导的采购意向。遗憾的是,医院正值预算紧张时期,最近由于雇佣了一位年轻有为的电气工程专业的毕业生而将预算进行了紧缩。为了满足放射学家的要求,院方向年轻的工程师布置了一项任务,让他设计一个系统(院方认为设计和构造一个相似的自用系统会较为便宜。医院现在拥有这样的系统的某些部件,但传输原始X光数据的时间要超过两分

钟)。院方要求工程师在下午的工作人员会议上拿出一个初始框图。由于时间紧且手头只有一本《数字图像处理》的课本(当然需要这本书),工程师能够在概念上设计一个系统以满足传输和相关的压缩要求。构建这个系统的概念框图,详细说明你将采用的压缩技术。

8.25 说明式(8.6.2)规定的基于提升的小波变换与使用表 8.20 中系数的传统 FWT 滤波器组是等价的。以 $\alpha,\beta,\gamma,\delta$ 和 K 的形式定义滤波器系数。

8.26 计算 JPEG 2000 编码图像的子带量化步长。图像采用隐式量化,对 $2LL$ 子带的尾数和指数分配 8 比特数据。

★8.27 为图 8.47 的编码器画出相应的 MPEG 解码器的框图。

第 9 章　形态学图像处理

形态学(morphology)一词通常代表生物学的一个分支。它是研究动物和植物的形态和结构的学科。我们在这里使用同一词语表示数学形态学的内容,将数学形态学(mathematical morphology)作为工具从图像中提取对于表达和描绘区域形状有用处的图像分量,比如边界、骨架以及凸壳,等等。我们对用于预处理或后处理的形态学技术同样感兴趣,比如形态学过滤、细化和修剪,等等。

数学形态学的语言是集合论。同样,形态学为大量的图像处理问题提供了一种一致的有力方法。数学形态学中的集合表示图像中的不同对象。例如,在二值图像中,所有黑色像素的集合是图像完整的形态学描述。在二值图像中,正被讨论的集合是二维整数空间(Z^2)的元素(见 2.4.2 节),在这个二维整数空间中,集合的每个元素都是一个多元组(二维向量),这些多元组的坐标是一个黑色(或白色,取决于事先的约定)像素在图像中的坐标(x,y)。灰度级数字图像可以表示为 Z 空间(Z^3)上分量的集合。在这种情况下,集合中每个元素的两个分量是像素的坐标,第 3 个分量对应于像素的离散灰度级值。更高维度空间中的集合可以包含图像的其他属性,比如,颜色和随时间变化的分量,等等。

在下面的章节中,我们将建立并说明几个在数学形态学中重要的概念。这里介绍的许多概念可在 n 维欧几里得空间(E 的 n 次方)中加以公式化。然而,我们的兴趣一开始是在二值图像上的,这种图像的各个分量是 Z^2 的元素。我们在 9.6 节中将讨论范围扩展到灰度级图像。

从本章的材料开始,我们关注的焦点将有所转变。以前,我们关注的是纯粹的图像处理方法,这种方法的输入和输出都是图像。而以后,我们关注的处理方法将变为输入是图像,而输出为从这些图像中提取出来的属性。在 1.1 节中对这方面进行了定义。像形态学及与其相关的概念这类工具是实现从图像中提取"内涵"这一目的所需的数学基础的基石。其他方法在本书的余下章节中将给予探讨和应用。

9.1　序言

本节中,我们介绍几个集合论的基本概念。集合理论是本章各小节的基础。

9.1.1　集合论的几个基本概念

令 A 为一个 Z 中的集合。如果 $a=(a_1,a_2)$是 A 的元素,则我们将其写成:

$$a \in A \tag{9.1.1}$$

同样,如果 a 不是 A 的元素,我们写成:

$$a \notin A \tag{9.1.2}$$

不包含任何元素的集合称为空集,用符号$\varnothing$表示。

集合由两个大括号之中的内容表示:$\{\cdot\}$。本章中我们关注的集合元素是,图像中描述的对象或其他感兴趣特征的像素坐标。例如,当我们写出形如 $C=\{w \mid w=-d, d\in D\}$的表达

式时，我们表达的意思是：集合 C 是元素 w 的集合。而 w 是通过用 -1 与集合 D 中的所有元素的两个坐标相乘得到的。

如果集合 A 的每个元素又是另一个集合 B 的一个元素，则 A 称为 B 的子集，表示为：

$$A \subseteq B \tag{9.1.3}$$

两个集合 A 和 B 的并集表示为：

$$C = A \cup B \tag{9.1.4}$$

这个集合包含集合 A 和 B 的所有元素。同样，两个集合 A 和 B 的交集表示为：

$$D = A \cap B \tag{9.1.5}$$

这个集合包含的元素同时属于集合 A 和 B。

如果 A 和 B 两个集合没有共同元素，则称为不相容的或互斥的。此时，

$$A \cap B = \varnothing \tag{9.1.6}$$

集合 A 的补集是不包含于集合 A 的所有元素组成的集合：

$$A^c = \{w \mid w \notin A\} \tag{9.1.7}$$

集合 A 和 B 的差，表示为 $A - B$，定义为：

$$A - B = \{w \mid w \in A, w \notin B\} = A \cap B^c \tag{9.1.8}$$

我们看出这个集合中的元素属于 A，而不属于 B。图 9.1 用图示说明了前述概念。集合运算的结果在图中用灰色表示。

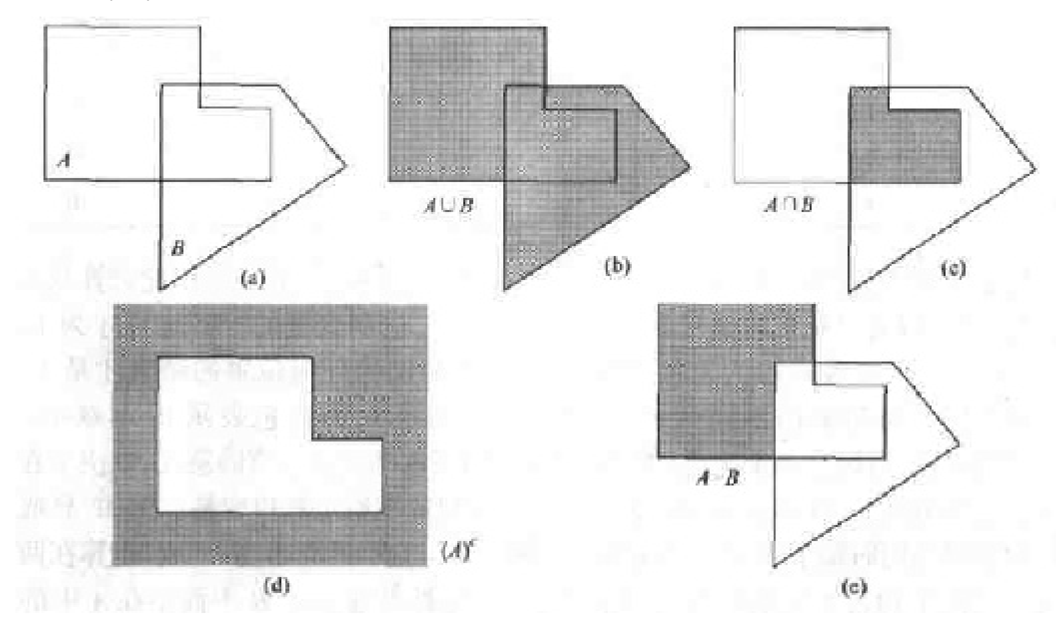


图 9.1　(a)两个集合 A 和 B，(b)A 和 B 的并集，(c)A 和 B 的交集，(d)A 的补集，(e)A 和 B 的差

我们还需要另外两个能广泛应用于形态学的附加定义，但通常在集合论的基本内容中无法找到。集合 B 的反射，表示为 $\hat{B}$，定义为：

$$\hat{B} = \{w \mid w = -b, b \in B\} \tag{9.1.9}$$

集合 A 平移到点 $z = (z_1, z_2)$，表示为 $(A)_z$，定义为：

$$(A)_z = \{c \mid c = a + z, a \in A\} \tag{9.1.10}$$

图 9.2 用来自图 9.1 的集合说明了这两个定义。黑点标识了图中集合的原点。

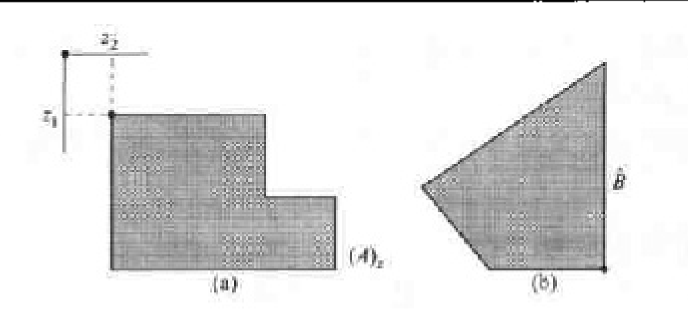


图 9.2 (a)集合 A 平移到 z,(b)集合 B 的反射,集合 A 和 B 来自图 9.1

9.1.2 二值图像的逻辑运算

很多应用是以本章讨论的形态学概念为基础的,并涉及二值图像。逻辑运算尽管本质上很简单,但对于实现以形态学为基础的图像处理算法是一种有力的补充手段。与掩模有关的逻辑运算在 3.4 节中介绍。在下面的讨论中,我们关注的是涉及二值像素和图像的逻辑运算。

在图像处理中用到的主要逻辑运算是:与、或和非(求补)。表 9.1 中总结了这些运算的性质。这些运算在功能上是完善的。它们可以互相组合形成其他逻辑运算。

表 9.1 三种基本的逻辑运算

p	q	p 与 q $(p\cdot q)$	p 或 q $(p+q)$	非$(p)(\bar{p})$
0	0	0	0	1
0	1	0	1	1
1	0	0	1	0
1	1	1	1	0

在两幅或多幅图像的对应像素间逐像素进行逻辑运算(除了“非”运算,此运算只对单一图像的像素进行)。因为只有在两个变量都是 1 时,两个二进制变量的“与”运算才为 1,所以只有两幅输入图像的对应像素均为 1 时,“与”运算的结果图像任何位置的结果才是 1。图 9.3 显示了涉及图像逻辑运算的不同例子。这里,黑色表示 1,而白色表示 0[本章中,我们使用两种约定的说法,有时会颠倒暗色(黑色或灰色)和亮色(白色)二值的意义,取决于在给定情况下哪种表达更清楚]。使用表 9.1 中的定义构造其他的逻辑运算很容易。例如,异或运算是当两个像素的值不同时结果为 1,否则为 0。这种运算与“或”运算不同。“或”运算在两个像素有一个为 1 或两个均为 1 时结果为 1。同样,非-与运算可选出在 B 中而不在 A 中的黑色像素。

注意到刚才讨论的逻辑运算与 9.1.1 节中讨论的集合运算有一一对应的关系是很重要的,并且逻辑运算被限制只对二进制变量进行运算,而这通常不是集合运算所处理的情况。因此,比如集合论中的交集运算在运算对象为二进制变量时归为“与”运算。像“相交”和“与”这类术语(甚至它们的符号)经常在各种著述中交替地用于表示一般的或二进制值的集合运算。通常从讨论的上下文中可以清楚地知道它们的意义。

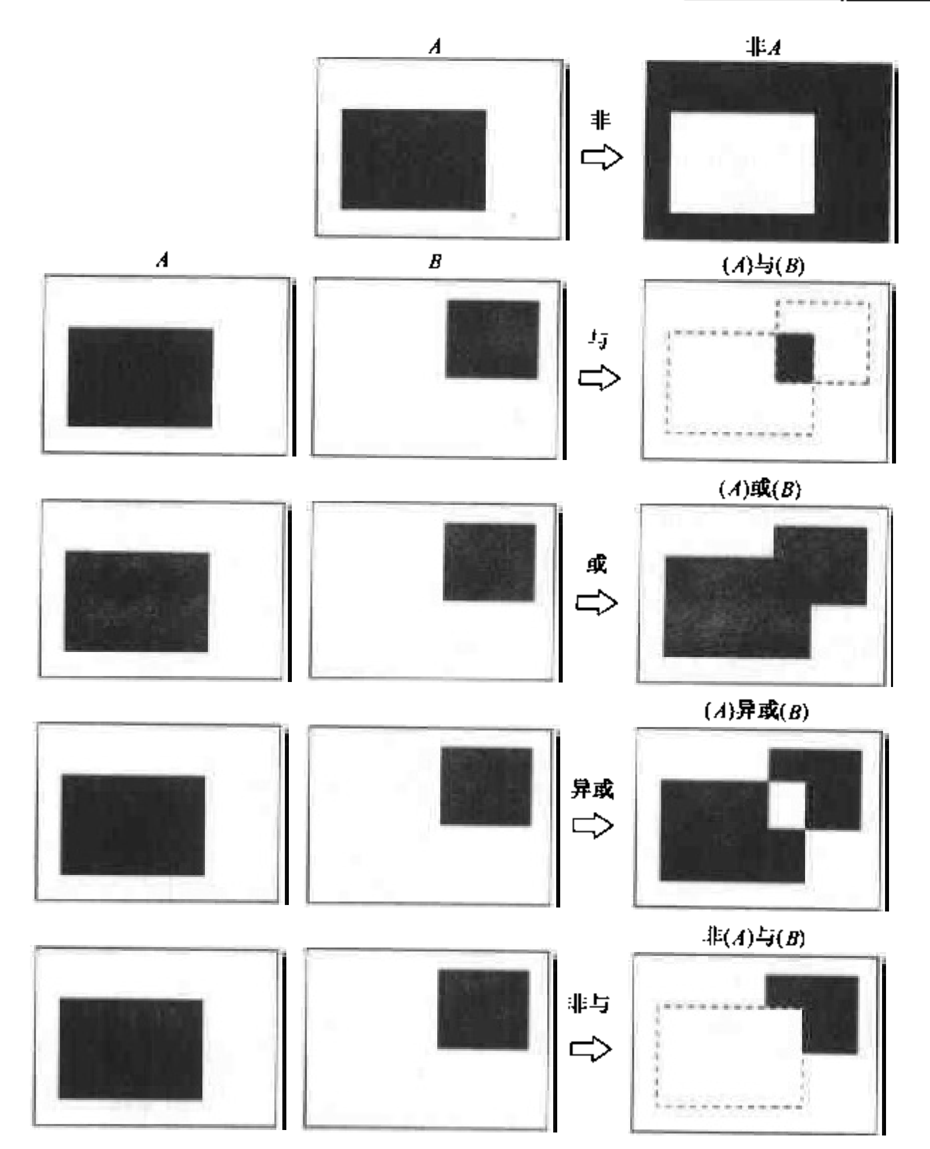


图 9.3　二值图像之间的一些逻辑运算。本例中，黑色表示二进制 1，白色表示二进制 0

9.2　膨胀与腐蚀

我们通过详细探讨两类操作——膨胀和腐蚀——开始我们对形态学上的操作的讨论。这两种操作是形态学处理的基础。实际上，本章中讨论的许多形态学算法都是以这两种原始运算为基础的。

9.2.1　膨胀

由于 A 和 B 是 Z^2 中的集合，A 被 B 膨胀定义为：

$$A \oplus B = \{z \mid (\hat{B})_z \cap A \neq \varnothing\} \tag{9.2.1}$$

这个公式是以得到 B 的相对于它自身原点的映象并且由 z 对映象进行位移为基础的。A 被 B 膨胀是所有位移 z 的集合，这样，$\hat{B}$ 和 A 至少有一个元素是重叠的。根据这种解释，式(9.2.1)可以重写为：

$$A \oplus B = \{z \mid [(\hat{B})_z \cap A] \subseteq A\} \tag{9.2.2}$$

与在其他形态学运算中一样,集合 B 通常叫做膨胀的结构元素。

在当前关于形态学方面的著述中式(9.2.1)并不是膨胀的惟一定义形式(见习题 9.10 和习题 9.11 中两种不同但等价的定义形式)。然而,当把结构元素 B 看做一个卷积模板时,式(9.2.1)的定义形式比起其他定义形式更为直观。这是它区别于其他定义形式的突出优点。尽管膨胀是以集合运算为基础,而卷积是以算术运算为基础,但相对于 B 的原点对 B 进行翻转,而后逐步移动 B 以便 B 能滑过集合(图像)A,这一基本过程与 3.5 节和 4.2.4 节中讨论的卷积过程是相似的。

图 9.4(a)显示了一个简单的集合,图 9.4(b)显示了一个结构元素和它的映像(黑色点表示元素的原点)。此时,结构元素和它的映像是相等的,因为 B 关于它的原点对称。图 9.4(c)中的虚线显示了作为基准的初始集合,实线显示了对 $\hat{B}$ 的原点进一步移动到 z 的限制。超出这个限制会使 $\hat{B}$ 和 A 的交集为空。所以,所有处在这一边界之内的点构成了使用 B 进行的 A 的膨胀。图 9.4(d)显示了一个被设计用来在垂直方向比水平方向进行更多膨胀的结构元素。图 9.4(e)显示了用这个结构元素进行膨胀的结果。

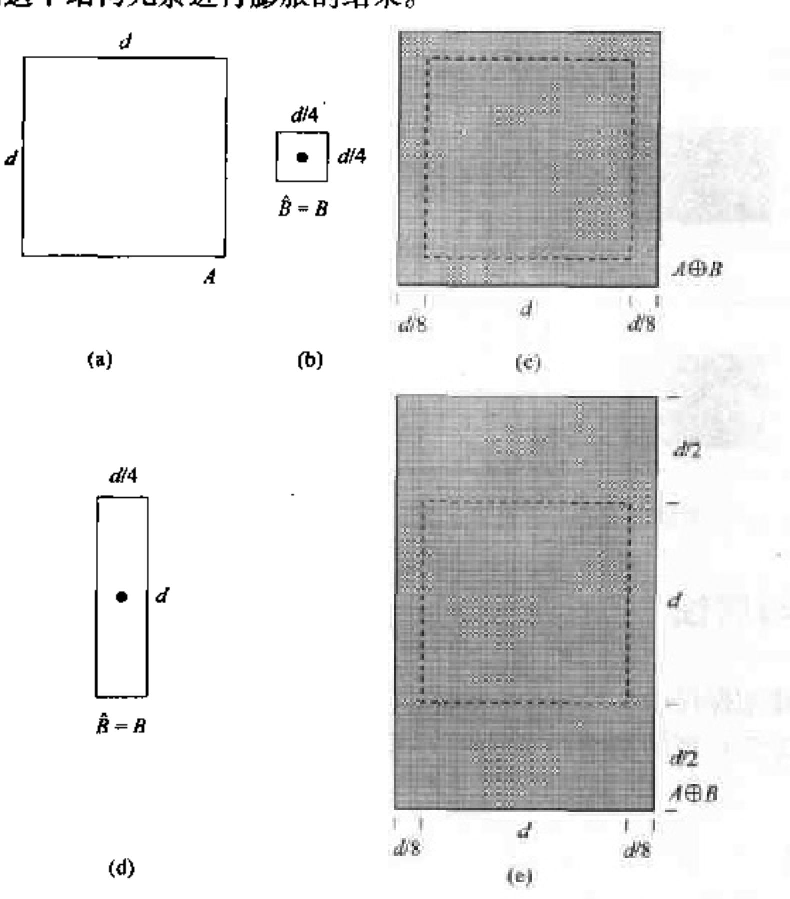


图 9.4 (a)集合 A,(b)方形结构元素(黑点为中心),(c)B 对 A 的膨胀以阴影显示,(d)拉长的结构元素,(e)使用这个元素进行的 A 的膨胀

例 9.1 将裂缝桥接起来的形态学膨胀的应用

膨胀最简单的应用之一是将裂缝桥接起来。图 9.5(a)中显示了与我们研究过的图 4.19 相同的带有间断的图像,这与低通滤波器的使用有关。已知间断的最大长度为两个像素。

图 9.5(b)显示了能够修复这些间断的简单结构元素。图 9.5(c)显示了使用这个结构元素对原图进行膨胀后的结果。现在间断已被连接起来了。形态学方法优于我们在图 4.19中用于连接间断的低通滤波方法的一个直接优点就是,这种方法在一幅二值图像中直接得到结果。而低通滤波方法从二值图像开始并生成一幅灰度级图像,这幅灰度级图像需要用门限函数进行一次处理才能将它转变回二值图像。

0	1	0
1	1	1
0	1	0

图 9.5　(a)带有间断字符的低分辨率示例文本(放大图),(b)结构元素,(c)通过(b)对(a)膨胀。断线被连接起来了

9.2.2 腐蚀

对 Z 中的集合 A 和 B,使用 B 对 A 进行腐蚀,用 $A \ominus B$ 表示,并定义为:

$$A \ominus B = \{z \mid (B)_z \subseteq A\} \tag{9.2.3}$$

一句话,这个公式说明,使用 B 对 A 进行腐蚀是所有 B 中包含于 A 中的点 z 的集合用 z 平移。同膨胀的情况一样,式(9.2.3)并不是腐蚀惟一的定义形式(见习题 9.12 和习题 9.13 的两种不同但等价的定义形式)。然而,由于与式(9.2.1)相同的原因,在形态学实际应用过程中人们偏爱使用式(9.2.3)。

图 9.6 显示了与图 9.4 相似的过程。对比图 9.6(c),以前的集合 A 显示为虚线。阴影区域的边界说明 B 的原点进一步移动的界限。超出这个界限会使集合不再完全包含于集合 A 中。因此,在这个边界内(也就是阴影区域),点的位置构成了使用 B 对 A 进行的腐蚀。图 9.6(d)显示了一个拉长的结构元素,图 9.6(e)显示了用这个元素腐蚀 A 的结果。注意原来的集合被腐蚀成一条线。

膨胀和腐蚀对于集合求补运算和反射运算是彼此对偶的。即:

$$(A \ominus B)^c = A^c \oplus \hat{B} \tag{9.2.4}$$

为了说明确定形态学表达式有效性的典型方法,我们对这个结果进行正规的证明。从腐蚀的

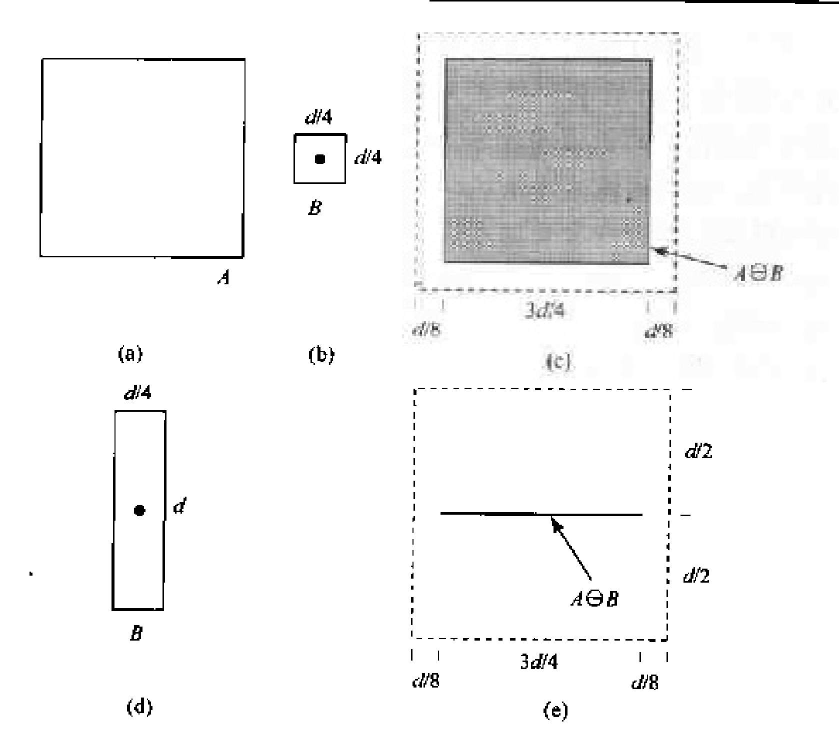


图 9.6 (a)集合 A,(b)方形结构元素,(c)由 B 对 A 腐蚀,如阴影所示,(d)拉长的结构元素,(e)使用这个元素对 A 腐蚀

定义开始,我们有:

$$(A \ominus B)^c = \{z \mid (B)_z \subseteq A\}^c$$

如果集合$(B)_z$包含于集合 A,则$(B)_z \cap A^c = \varnothing$,此时前述公式变为:

$$(A \ominus B)^c = \{z \mid (B)_z \cap A^c = \varnothing\}^c$$

但,满足$(B)_z \cap A^c = \varnothing$的 z 的集合的补集是满足$(B)_z \cap A^c \neq \varnothing$的集合。因此:

$$(A \ominus B)^c = \{z \mid (B)_z \cap A^c \neq \varnothing\} = A^c \oplus \hat{B}$$

这是从式(9.2.1)得出的最后一步,这就证明了这个结论。

例 9.2 使用形态学腐蚀除去图像的某些部分

腐蚀的一种最简单的用途是从二值图像中消除不相关的细节(根据尺寸)。图 9.7(a)显示的二值图像包含边长为 1,3,5,7,9 和 15 个像素的正方形。假设这里只留下最大的正方形而除去其他的正方形,我们可以通过用比我们要保留的对象稍小的结构元素对图像进行腐蚀。在这个例子中我们选择 13×13 像素大小的结构元素。图 9.7(b)显示了用这个结构元素对原图像进行腐蚀后得到的结果。图中只有部分最大的正方形保留下来。如图 9.7(c)所示,我们可以通过使用用来腐蚀的结构元素对这 3 个正方形进行膨胀恢复它们原来 15×15 像素的尺寸(通常膨胀不能完全恢复被腐蚀的对象;见习题 9.18)。注意例子中所有 3 幅图像中的对象都是用白色像素表示,而不是像前面的例子中用黑色表示。如我们前面提到的,两种表达方法在实际中均有使用。除非申明,否则我们令结构元素中起作用的元素选择与感兴趣对象一样的二进制值。这个例子中提到的概念是形态学过滤的基础,将在接下来的小节中进行介绍。

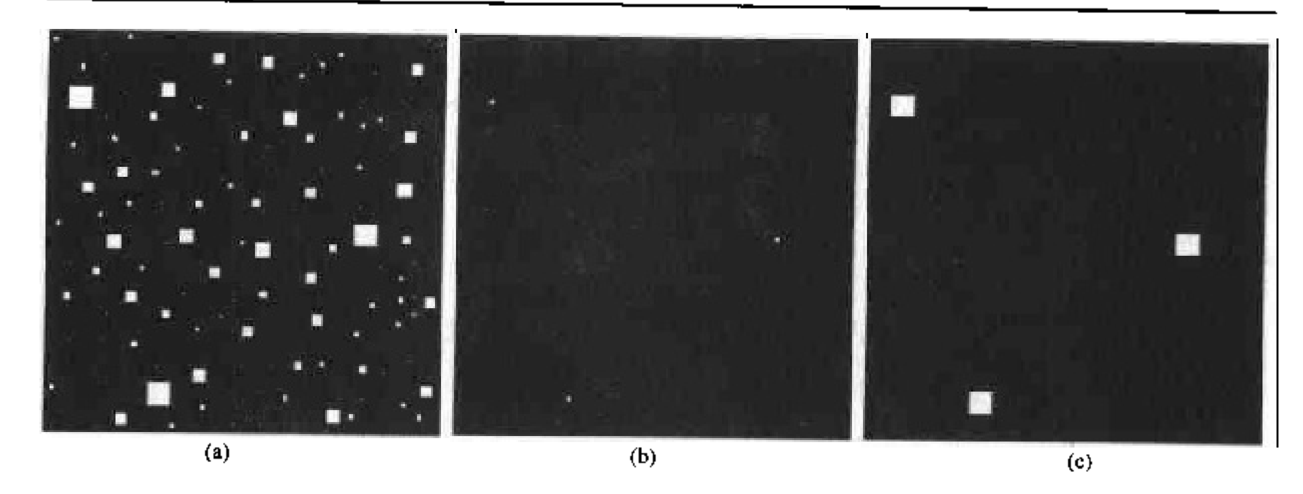


图 9.7　(a)图像内部边长为 1,3,5,7,9 和 15 个像素的正方形的图像,(b)用第一个方形结构元素(一边有 13 个像素)对(a)进行腐蚀,(c)使用相同的结构元素对(b)进行膨胀

9.3　开操作与闭操作

如我们所见到的,膨胀使图像扩大而腐蚀使图像缩小。本节中,我们讨论另外两个重要的形态学操作:开操作与闭操作。开操作一般使对象的轮廓变得光滑,断开狭窄的间断和消除细的突出物。闭操作同样使轮廓线更为光滑,但与开操作相反的是,它通常消弥狭窄的间断和长细的鸿沟,消除小的孔洞,并填补轮廓线中的断裂。

使用结构元素 B 对集合 A 进行开操作,表示为 $A\circ B$,定义为:

$$A \circ B = (A \ominus B) \oplus B \tag{9.3.1}$$

因此,用 B 对 A 进行开操作就是用 B 对 A 腐蚀,然后用 B 对结果进行膨胀。

同样,使用结构元素 B 对集合 A 的闭操作,表示为 $A\cdot B$,定义如下:

$$A \cdot B = (A \oplus B) \ominus B \tag{9.3.2}$$

这个公式说明,使用结构元素 B 对集合 A 的闭操作就是用 B 对 A 进行膨胀,而后用 B 对结果进行腐蚀。

开操作有一个简单的几何解释(图 9.8)。假设我们将结构元素 B 看做一个(扁平的)"转球"。$A\circ B$ 的边界通过 B 中的点完成,即 B 在 A 的边界内转动时,B 中的点所能到达的 A 的边界的最远点。这个开操作的几何拟合特性使我们得出了集合论的一个公式。这个公式说明用 B 对 A 进行开操作是通过求取 B 在拟合 A 时的平移的并集得到的。就是说,开操作可以表示为一个拟合操作:

$$A \circ B = \bigcup \{(B)_z \mid (B)_z \subseteq A\} \tag{9.3.3}$$

这里 $\bigcup\{\cdot\}$ 表示大括号中所有集合的并集。

闭操作有相似的几何解释,只是我们现在在边界的外部转动 B(图 9.9)。简而言之,开操作和闭操作是一对对偶操作,所以闭操作在边界外部转动球是预料之中的事。从几何上讲,当且仅当对包含 w 的 $(B)_z$ 进行的所有平移都满足 $(B)_z \cap A \neq \varnothing$ 时,点 w 是 $A\cdot B$ 的一个元素。图 9.9 说明了闭操作这一基本的几何性质。

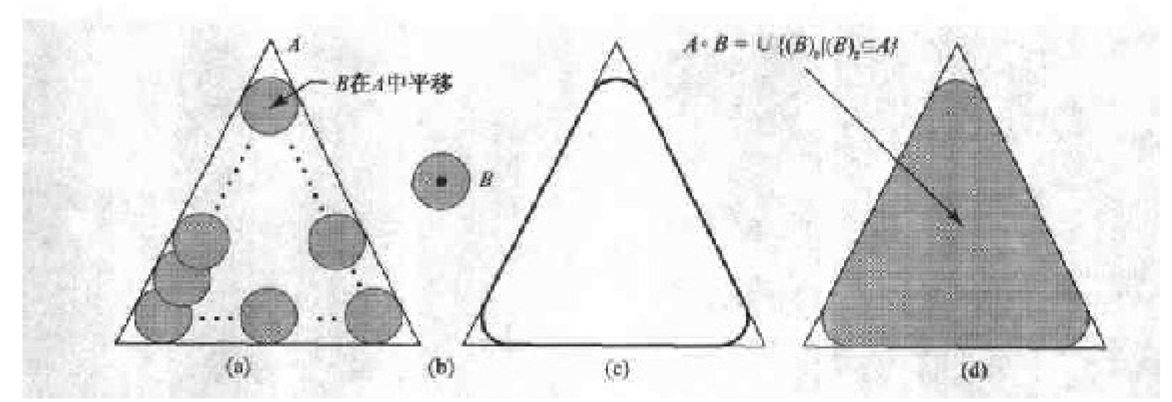


图 9.8 (a)结构元素 B 沿着 A 的内部边界转动(点表示 B 的圆心),(b)结构元素 B,(c)粗线是开操作的外部边界,(d)完全开操作(阴影部分)

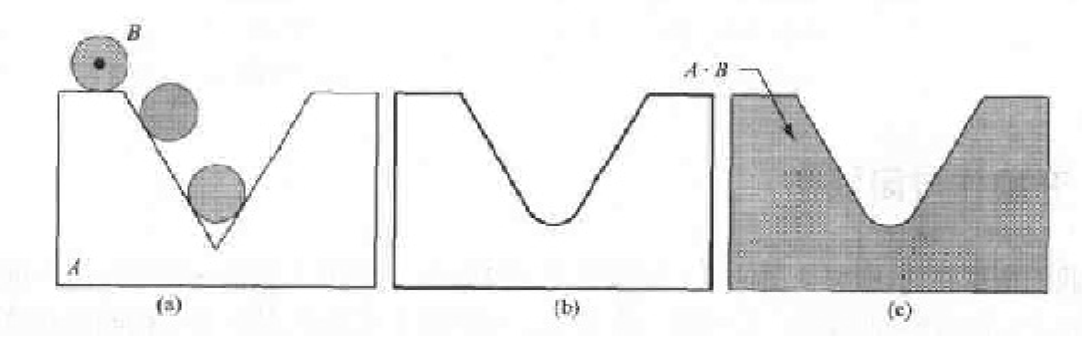


图 9.9 (a)结构元素 B 在集合 A 的外部边界上转动,(b)粗线表示闭操作的外部边界,(c)完全的闭操作(阴影部分)

例 9.3 对形态学上的开操作和闭操作的简单说明

图 9.10 进一步对开操作和闭操作进行了说明。图 9.10(a)显示了集合 A,图 9.10(b)显示了在腐蚀过程中的一块圆盘形结构元素的各种位置。当腐蚀完成时,得到图 9.10(c)中显示的连接图。注意,两个主要部分之间的桥接部分消失了。这部分的宽度与结构元素的直径相比较要细;就是说,这部分集合不能完全包含结构元素,因此,无法满足式(9.2.3)的条件。对象最右边的两个部分也是如此。圆盘无法拟合的突出部分被消除掉了。在图 9.10(d)中显示了对腐蚀后的集合进行膨胀的操作,图 9.10(e)显示了开操作最后的结果。注意,方向向外的角变圆滑了,而方向向内的角没受影响。

同样,图 9.10(f)到(i)显示了使用同样的结构元素对 A 进行闭操作的结果。我们注意到方向向内的拐角变得圆滑了,而方向向外的拐角没有变化。在 A 的边界上,最左边的侵入部分在尺寸上明显地减小了。因为在这个位置上,圆盘无法拟合。同时也要注意在使用圆形结构元素对集合 A 进行开操作和闭操作后,所得对象的各个部分得到了平滑处理。

如此处的膨胀和腐蚀一样,开操作和闭操作也是一对关于集合求补及映像的对偶操作。即:

$$(A \cdot B)^c = (A^c \circ \hat{B}) \tag{9.3.4}$$

我们将这个结果的证明作为练习(习题 9.14)留给读者。

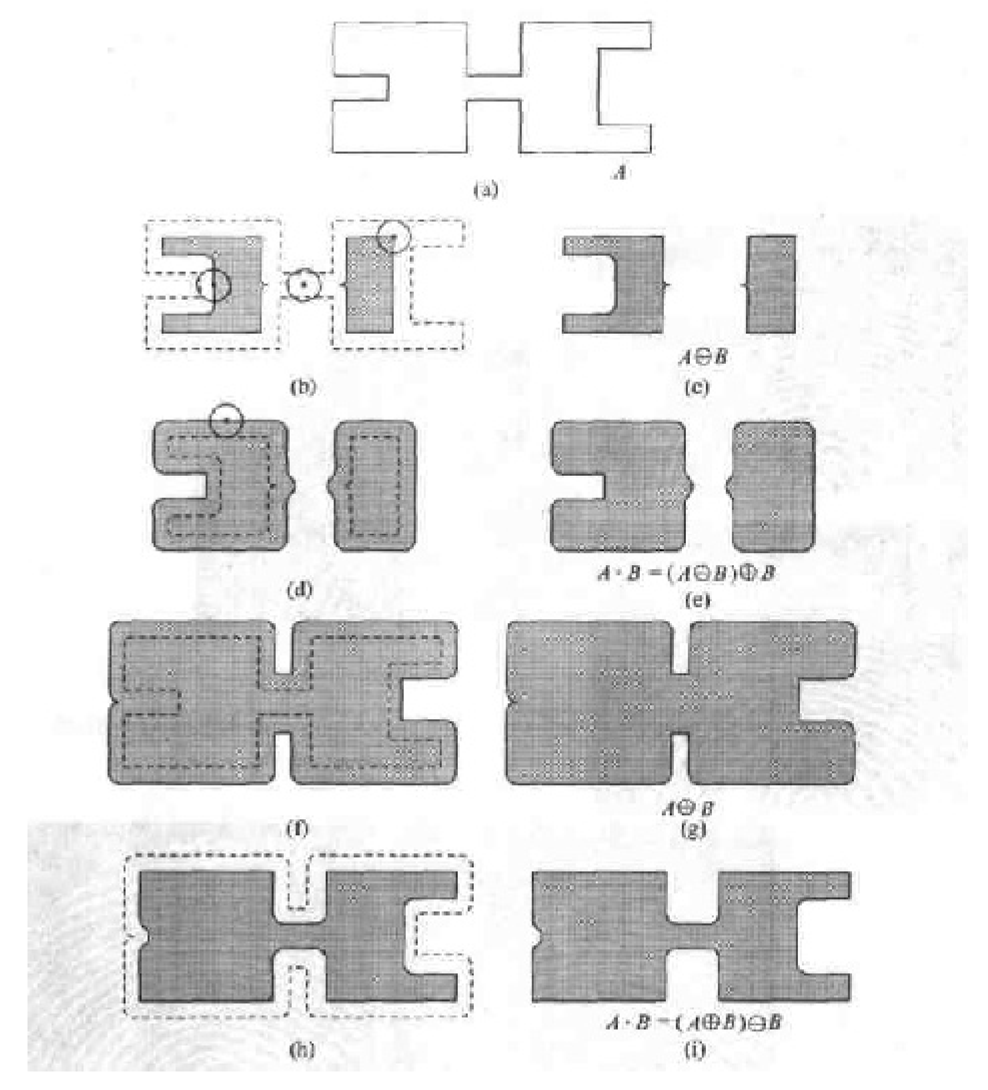


图 9.10　形态学开操作和闭操作。结构元素是(b)中各部分显示的小圆。黑点是结构元素的中心

开操作满足下列性质:

(i) $A\circ B$ 是 A 的子集合(子图)。
(ii) 如果 C 是 D 的子集,则 $C\circ B$ 是 $D\circ B$ 的子集。
(iii) $(A\circ B)\circ B=A\circ B$。

同样,闭操作也满足下列性质:

(i) A 是 $A\cdot B$ 的子集(子图)。
(ii) 如果 C 是 D 的子集,则 $C\cdot B$ 是 $D\cdot B$ 的子集。
(iii) $(A\cdot B)\cdot B=A\cdot B$。

注意,由两种情况下的条件(iii)可知,算子应用一次后,一个集合进行多少次开操作或闭操作都不会有变化。

例 9.4　形态学滤波的开操作和闭操作的使用

形态学运算可以用于构造与第 3 章中讨论的空间滤波概念相类似的滤波器。显示于

图 9.11(a)中的二值图像显示了受噪声污染的部分指纹图像。这里噪声表现为黑色背景上的亮元素和亮指纹部分的暗元素。我们的目的是消除噪声及它对印刷所造成的影响，使图像失真尽可能减小。由闭操作后紧跟着进行开操作形成的形态学滤波器可用于实现这个目的。

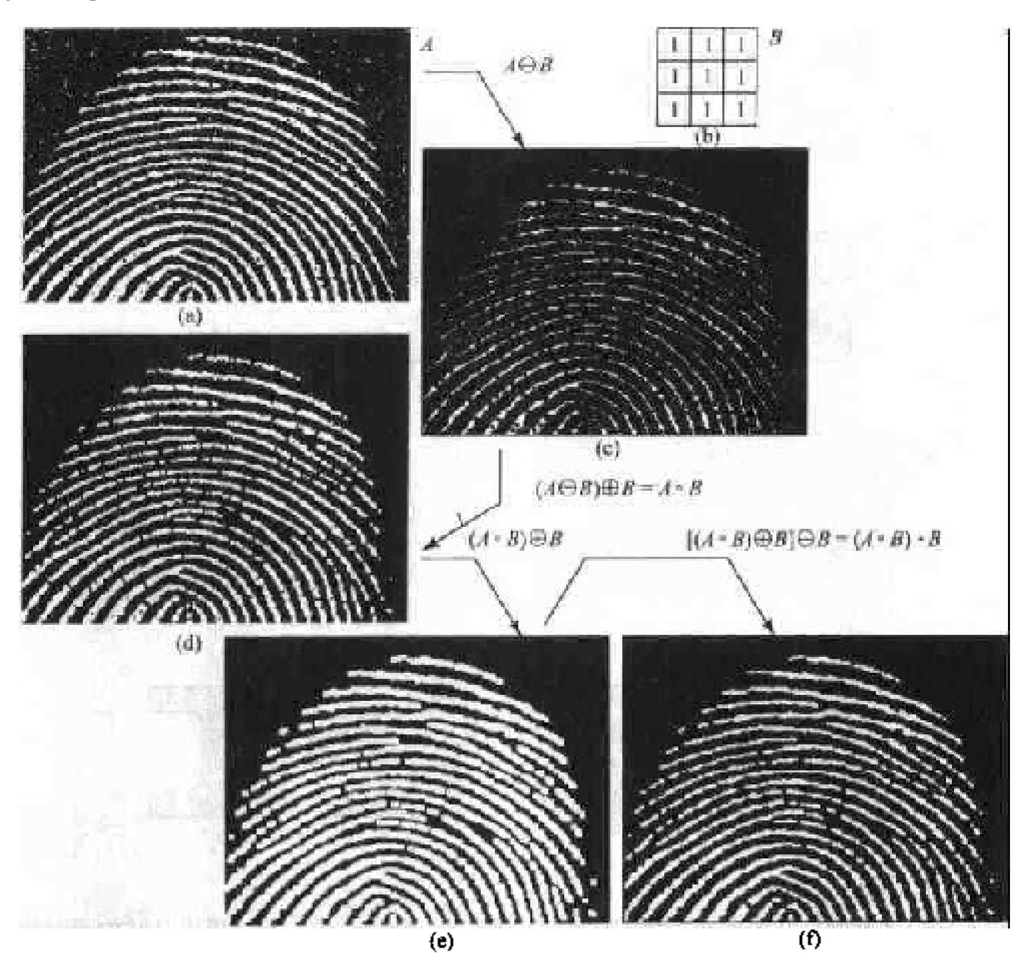


图 9.11　(a)有噪声的图像，(c)腐蚀图像，(d)A 的开操作，(e)开操作的膨胀，(f)开操作的闭操作[这个例子的原图由美国国家标准技术研究所(NIST)提供]

图 9.11(b)显示了所使用的结构元素。图 9.11 余下的部分显示了滤波操作的每一步过程。图 9.11(c)显示了使用结构元素对 A 进行腐蚀的结果。背景噪声在开操作的腐蚀过程中被完全消除了，因为在这种情况下，图像中噪声部分的物理尺寸均比结构元素小。而包含于指纹中的噪声元素(黑点)的尺寸却有增加。原因是，当目标被腐蚀时，这些元素被作为应该增大尺寸的内部边界进行了处理。这种增大在图 9.11(c)中进行膨胀的过程中被抵消了。图 9.11(d)显示了该结果。包含于指纹中的噪声分量的尺寸被减小或被完全消除掉了。

刚才描述的两种操作构成了用 B 对 A 进行的开操作。我们注意到图 9.11(d)中开操作的实际效果是消除背景和指纹中的所有噪声。然而，在指纹纹路间产生了新的间断。为了防止这种不希望的影响，我们在开操作的基础上进行膨胀，如图 9.11(e)所示。大部分间断被恢复了，但指纹的纹路变粗了，可以通过腐蚀来弥补出现的这种情况。示于图 9.11(f)的结果构成了对图 9.11(d)中开操作的闭操作。最后结果的噪声斑点清除得相

当干净，但这种方法有缺点，即有些指纹纹路没有被完全修复，并还有间断。对这种情况的改进也并非毫无希望，只是因为这一过程中在保持连续性方面没有提出任何条件而已（我们将在例 9.8 中讨论这一问题并在 11.1.5 节中给出处理这个问题的方法）。

9.4 击中或击不中变换

形态学上的击中或击不中变换是形状检测的基本工具。我们用图 9.12 作为辅助介绍这个概念。图 9.12 显示了一个由 3 种形状（子集）组成的集合 A，子集用 X，Y 和 Z 表示。图 9.12(a)到(c)中的阴影部分指明了初始集合，而图 9.12(d)和(e)中的阴影部分指出了进行形态学操作后的结果。目的是找到 3 种形状之一的位置，如 X 的位置。

令每种形状的重心为它的原点。设 X 被包围在一个小窗口 W 中。与 W 有关的 X 的局部背景定义为集合的差($W-X$)，如图 9.12(b)所示。图 9.12(c)显示了 A 的补集，在后面将使用到它。图 9.12(d)显示了由 X 对 A 腐蚀的结果（显示虚线作为参考）。使用 X 对 A 进行的腐蚀是 X 原点位置的集合。这样，X 就完全包含在 A 中了。换一个角度解释，$A\ominus X$ 从几何上可以被看做 X 的原点所有位置的集合，在这些位置 X 找到了在 A 中的匹配（击中）。请记住图 9.12 中 A 只包含 3 种彼此不相连的集合 X，Y 和 Z。

图 9.12(e)显示了由局部背景集合($W-X$)对集合 A 的补集腐蚀的结果。图 9.12(e)的外圈阴影区域是腐蚀部分。我们根据图 9.12(d)和(e)注意到，X 在 A 内能得到精确拟合的位置集合，是由 X 对 A 的腐蚀和由($W-X$)对 A^c 的腐蚀的交集，如图 9.12(f)所示。这个交集正好是我们要找的位置。换句话说，如果 B 表示由 X 和 X 的背景构成的集合，则在 A 中对 B 进行的匹配（或匹配操作的集合）表示为 $A\circledast B$：

$$A\circledast B=(A\ominus X)\cap[A^c\ominus(W-X)] \tag{9.4.1}$$

我们可以通过令 $B=(B_1,B_2)$ 对这种表示法稍微进行推广。这里 B_1 是由与一个对象相联系的 B 元素构成的集合。B_2 是与相应背景有关的 B 元素的集合。根据前面的讨论，$B_1=X$，$B_2=(W-X)$。用这个表示方法，式(9.4.1)变为：

$$A\circledast B=(A\ominus B_1)\cap(A^c\ominus B_2) \tag{9.4.2}$$

因此，集合 $A\circledast B$ 同时包含了所有的原点，B_1 在 A 内找到匹配，B_2 在 A^c 中找到匹配。通过应用式(9.1.8)给出的集合之差的定义和式(9.2.4)给出的腐蚀与膨胀间的对偶关系，我们可以将式(9.4.2)写成：

$$A\circledast B=(A\ominus B_1)-(A\oplus\hat{B}_2) \tag{9.4.3}$$

然而，式(9.4.2)更为直观。我们将上述 3 个公式称为形态学上的击中或击不中变换。

使用与对象有关的结构元素 B_1 和与背景有关的 B_2 的原因是基于以下假设的定义，即，只有在两个或更多对象构成彼此不相交（不连通）的集合时，这些对象才是可区分的。要保证这个假设，需要在每个对象周围至少被一圈一个像素宽的背景围绕的条件。在某些应用中，我们也许对在某个集合中检测 1 和 0 组成的某种模式感兴趣，而此时是不需要背景的。在这种场合，击中或击不中变换转变成简单的腐蚀过程。正如前边所指出的那样，腐蚀是进行一系列的匹配，但对于检测单个对象来说不需要额外的背景匹配。这种简化的模式检测方案将用于下面章节讨论的某些算法中。

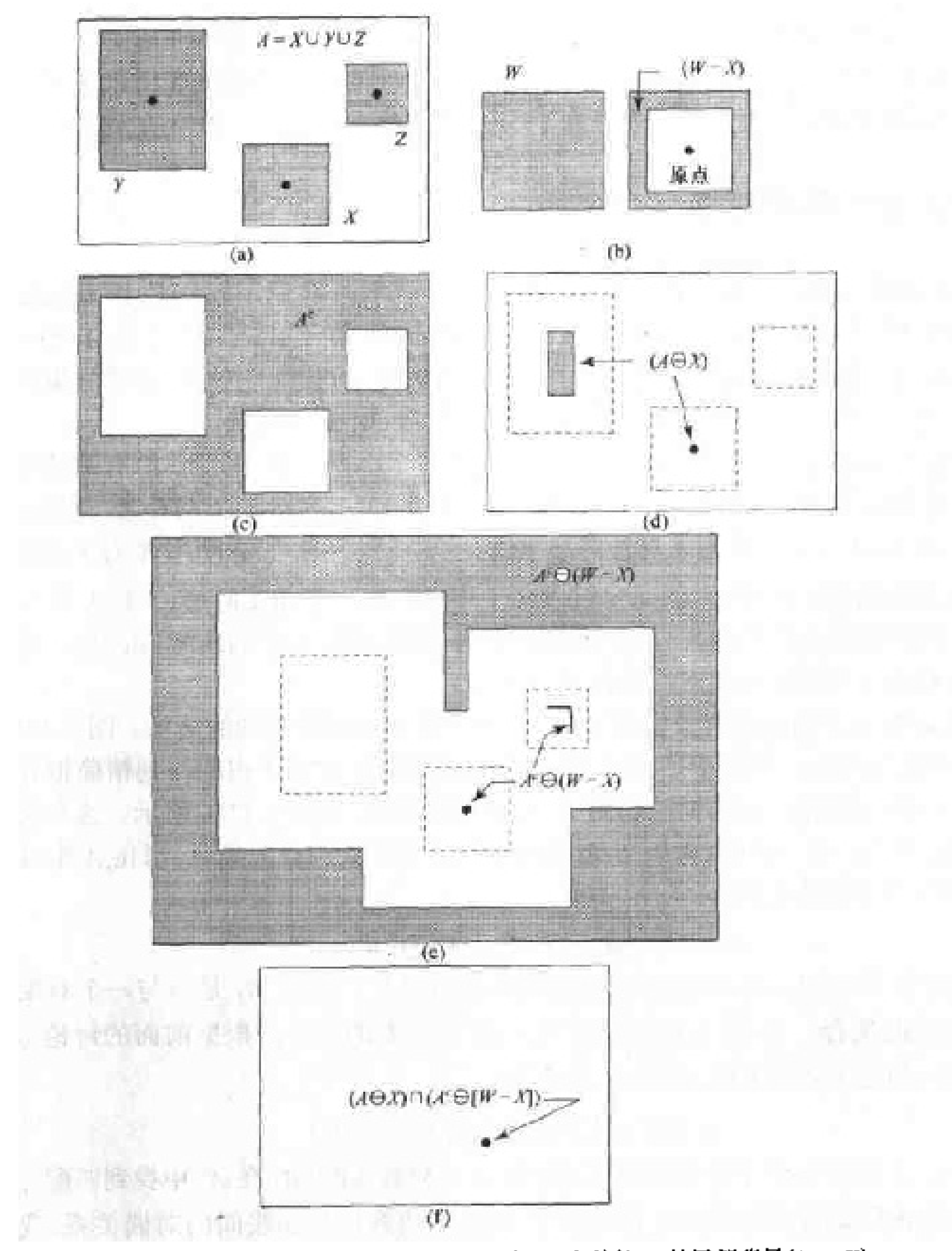


图 9.12 (a)集合 A,(b)窗口 W 和与 W 有关的 X 的局部背景($W-X$),(c) A 的补集,(d)用 X 对 A 腐蚀,(e)用($W-X$)对 A 腐蚀,(f)(d)和(e)的交集,显示了我们希望得到的 X 的原点位置

9.5 一些基本的形态学算法

以前面的讨论作为背景,我们现在可以考虑一些形态学的实际用途。当处理二值图像时,形态学的主要应用是提取对于描绘和表达形状有用的图像成分。我们特别要考虑提取边界、连通分量、凸壳和区域骨架的形态学算法。我们也要探讨几种经常与这些算法有联系的预处理或后处理方法(用于区域填充、细化、粗化和修剪)。在我们介绍每一种形态学处理时,

为了弄清楚每种处理的机理,将在这一小节中广泛使用"小图像"。图像是二值的,用 1 显示阴影区域,而 0 显示白色。

9.5.1　边界提取

集合 A 的边界表示为 $\beta(A)$,它可以通过先由 B 对 A 腐蚀,而后用 A 减去腐蚀得到。即:

$$\beta(A) = A - (A \ominus B) \tag{9.5.1}$$

这里 B 是一个适当的结构元素。

图 9.13 说明了边界提取的机理。这幅图像显示了一个简单的二值对象、一个结构元素 B 和使用式(9.5.1)的结果。

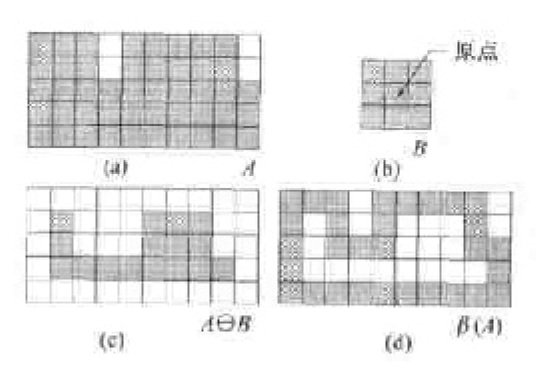


图 9.13　(a)集合 A,(b)结构元素 B,(c)使用 B 对 A 进行腐蚀,(d)由 A 减去腐蚀的结果得到边界

尽管图 9.13(b)中显示的结构元素是最常用的结构元素之一,但它绝对不是惟一的。例如,使用由 1 组成的 5×5 大小的结构元素将得到 2 到 3 个像素宽的边界。注意,当 B 的原点位于集合的边线上时,结构元素的一部分将处在图像的外面。对这种情况的一般处理方法是假设处于图像边界外部部分的值为 0。

例 9.5　使用形态学处理提取边界

图 9.14 对使用式(9.5.1)和图 9.13(b)中的结构元素进行了进一步的说明。在这个例子中,二进制值 1 表示为白色,0 表示为黑色,因此结构元的元素被处理成白色。因为结构元素的使用,示于图 9.14(b)中的边界是一个像素宽的。

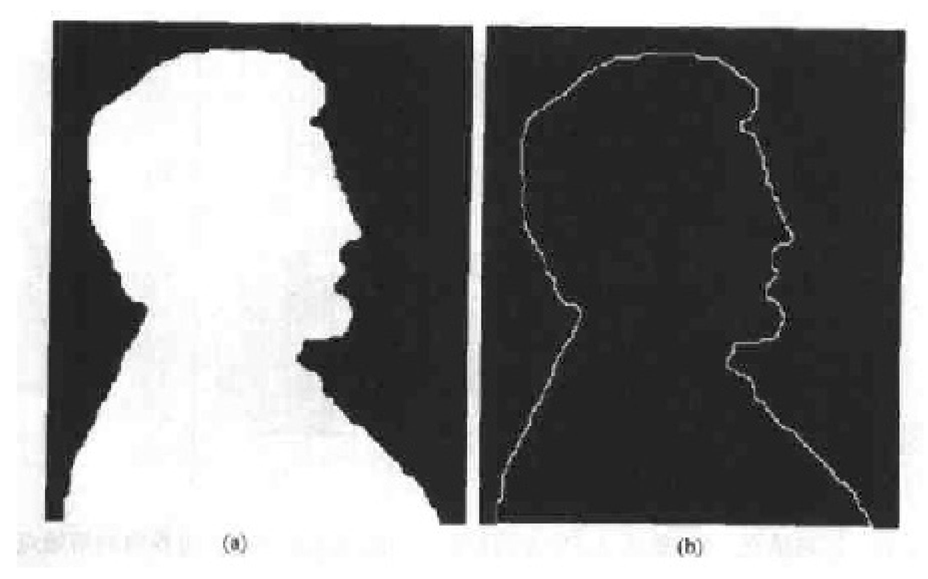


图 9.14　(a)一幅简单的二值图像,用 1 表示白,(b)使用式(9.5.1)和图 9.13(b)中的结构元素进行处理的结果

9.5.2 区域填充

接下来,我们探讨一个简单的用于区域填充的算法,它以集合的膨胀、求补和交集为基础。在图9.15中,A 表示一个包含子集的集合,其子集的元素均是区域的8连通边界点。目的是从边界内的一个点开始,用1填充整个区域。

如果我们采用惯例:所有非边界(背景)点标记为0,则以将1赋给 p 点开始。下列过程将整个区域用1填充:

$$X_k = (X_{k-1} \oplus B) \cap A^c \quad k = 1,2,3,\cdots \tag{9.5.2}$$

这里 $X_0 = p$,B 是示于图9.15(c)中的对称结构元素。如果 $X_k = X_{k-1}$,则算法在迭代的第 k 步结束。X_k 和 A 的并集包含被填充的集合和它的边界。

如果对公式的左部不加限制,则式(9.5.2)的膨胀处理将填充整个区域。但在每一步中,用 A^c 的交集将得到的结果限制在感兴趣区域内。这是我们关于形态学处理如何达到所要求特性的第一个例子。在这个应用中,上述处理被称为条件膨胀。图9.15余下部分进一步说明了式(9.5.2)所应用的技巧。尽管这个例子仅有一个子集,但如果假设在每条边界内都有一个给定点,则用于有限个这样的子集的概念是很清楚的。

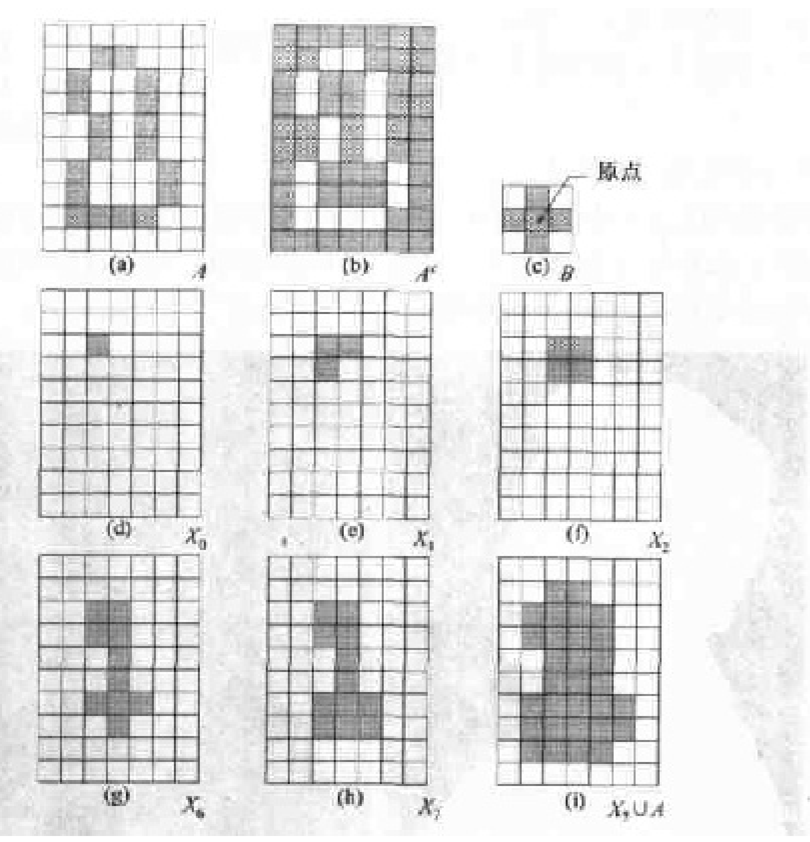


图9.15 区域填充。(a)集合 A,(b)A 的补集,(c)结构元素 B,(d)边界内的初始点,(e)~(h)式(9.5.2)使用的各个步骤,(i)最后的结果[(a)和(h)的并集]

例9.6 形态学区域填充

图9.16(a)显示了一幅由白色圆圈和其内部的黑色点组成的图像。这样的图像可以通过将包含磨光的球体(如滚珠)的场景用门限处理分为两个层次而得到。球内部的暗点是反

射的结果。我们的目的是通过区域填充消除反射。图9.16(a)显示了在球体中选择的一个点，图9.16(b)显示了将这一部分进行填充的结果。最后，图9.16(c)显示了填充所有球体后的结果。因为黑色点是背景点还是球体内部点必须是已知的，所以这个过程要完全自动化需要在算法中附加"智能"(见习题9.23)。

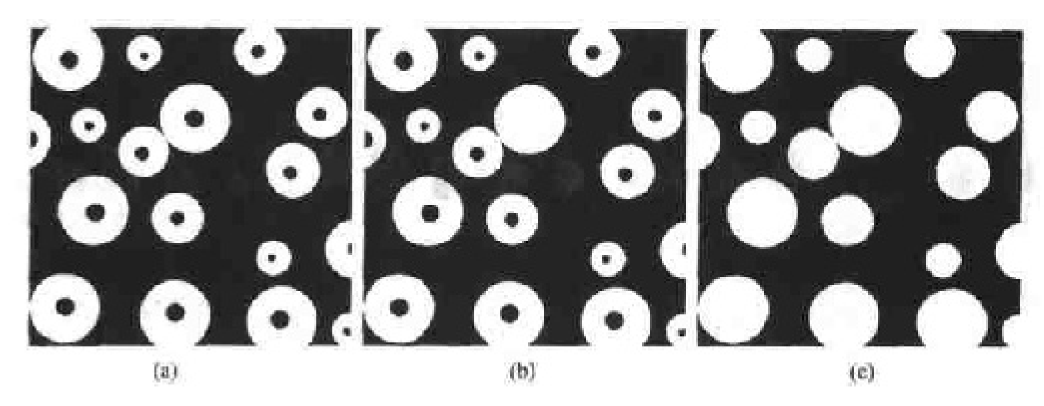


图9.16　(a)二值图像(区域内部的白色点表示区域填充算法的起点)，(b)区域填充的结果，(c)填充所有区域的结果

9.5.3　连通分量的提取

连通性和连通分量的概念在2.5.2节中介绍过了。实际上，在二值图像中提取连通分量是许多自动图像分析应用中的核心任务。令 Y 表示一个包含于集合 A 中的连通分量，并假设 Y 中的一个点 p 是已知的。而后，用下列的迭代表达式生成 Y 的所有元素：

$$X_k = (X_{k-1} \oplus B) \cap A \quad k = 1,2,3\cdots \tag{9.5.3}$$

这里 $X_0 = p$，B 是一个适当的结构元素，如图9.17所示。如果 $X_k = X_{k-1}$，算法收敛，并且我们令 $Y = X_k$。

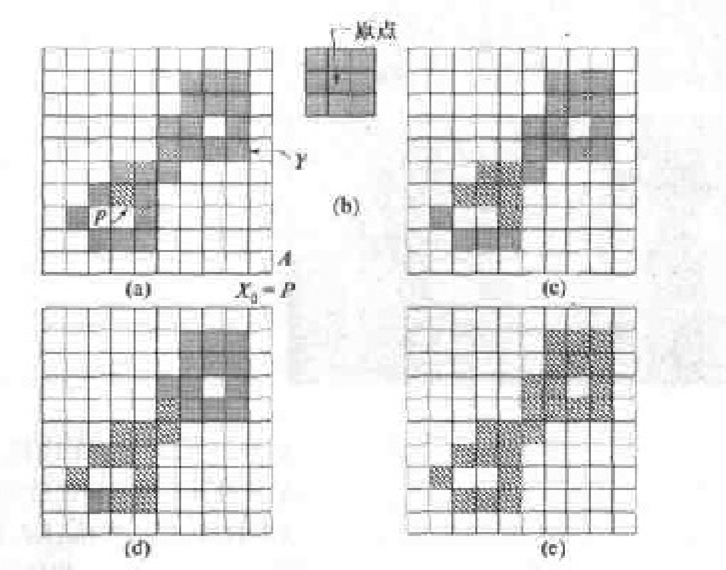


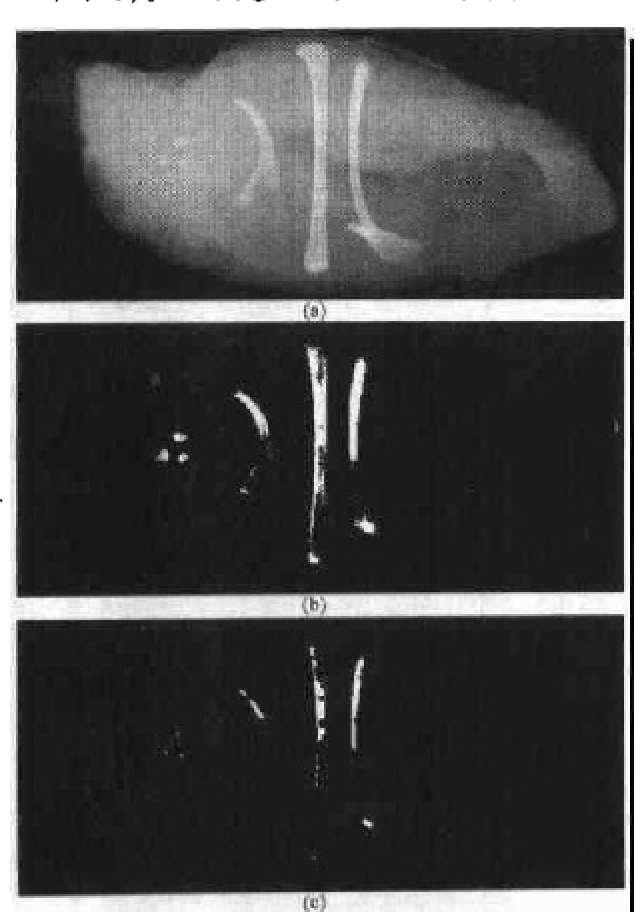
图9.17　(a)显示了起始点 p 的集合 A(所有阴影点值为1，但与 p 的表示不同，以说明这些点还没有被算法找到)，(b)结构元素，(c)第1次迭代的结果，(d)第2步的结果，(e)最终结果

式(9.5.3)在形式上与式(9.5.2)相似。仅有的差别是使用 A 代替了它的补集。此处的差别是由于寻找的所有元素(即,连通分量的元素)都被标记为了 1。

在每一步迭代操作中,与 A 的交集消除了位于中心的标记为 0 的元素膨胀。图 9.17 说明了式(9.5.3)所使用的技巧。注意,结构元素的形状假定在像素间具有 8 连通性。正如在区域填充算法中那样,如果假设每个连通分量中一个点已知,刚才讨论的结果适用于包含于 A 中的任何有限连通分量的集合。

例 9.7　使用连通分量检测包装食物中的外来物

连通分量经常用于自动监测。图 9.18(a)显示了一幅含有碎骨的鸡胸 X 光图像。将经过处理的食物在包装或运送之前进行外来物检测是很有意义的。在这个特定情况下,骨质的密度使它们在正常情况下的灰度级值与背景不同。这使通过一个单一门限将骨头从背景中提取出来变得很简单(在 3.1 节中介绍了门限处理。在 10.3 节中对此将进行相当详细的讨论)。结果是显示在图 9.18(b)中的二值图像。



连通分量	连通分量中像素的数目
01	11
02	9
03	9
04	39
05	133
06	1
07	1
08	743
09	7
10	11
11	11
12	9
13	9
14	674
15	85

(d)

图 9.18　(a)带有碎骨的鸡肉 X 光图像,(b)经门限处理的图像,(c)用元素为 1 的 5×5 大小的结构元素进行腐蚀后的图像,(d)(c)的连通分量中像素的数目(由德国 Diepholz 的 NTB Elektronische Geraete GmbH,www.ntbxray.com 提供图像)

在这幅图中最显著的特征是保留下来的点集聚为对象(骨头),而不是彼此孤立毫无关系的点。我们可以确定,只有具有"有效"尺寸的物体才能在对经门限处理的图像进行腐蚀

后保留下来。在这个例子中，我们将经过元素为 1 的 5×5 大小的结构元素腐蚀后保留下来的对象定义为有效对象。图 9.18(c)中显示了腐蚀的结果。下一步是分析保留下来的对象尺寸。我们通过在图像中提取连通分量来标记(识别)这些对象。图 9.18(d)中的表格列出了提取的结果。总共有 15 个连通分量，其中 4 个具有最大尺寸。这足以判断包含在原图中的明显外来物。如果需要，使用第 11 章中讨论的技术可以对对象做进一步的描述(比如对象的形状)。

9.5.4 凸壳

如果连接集合 A 内任意两个点的直线段都在 A 的内部，就称 A 是凸形的。任意集合 S 的凸壳 H 是包含 S 的最小凸集合。集合差 $H-S$ 称为 S 的凸缺。如在 11.1.4 节和 11.3.2 节中讨论的细节一样，凸壳和凸缺对于对象描绘是很有用处的。这里，我们介绍一种用于求取集合 A 的凸壳 $C(A)$ 的简单形态学算法。

令 $B^i, i=1,2,3,4$，表示图 9.19(a)中的 4 个结构元素。这个过程由计算公式组成：

$$X_k^i = (X_{k-1} \circledast B^i) \cup A \quad i=1,2,3,4 \quad k=1,2,3,\cdots \tag{9.5.4}$$

$X_0^i = A$。现在令 $D^i = X_{\text{conv}}^i$，这里下标“conv”表示在 $X_k^i = X_{k-1}^i$ 时收敛。A 的凸壳为：

$$C(A) = \bigcup_{i=1}^{4} D^i \tag{9.5.5}$$

换句话说，这一过程包括对 A 用 B^1 反复应用击中或击不中变换；当不再发生进一步的变化时，我们执行与 A 的并集运算，用 D^1 表示结果。这一过程用 B^2(用于 A)重复，直到不发生进一步的变化，如此反复。得到的 4 个 D 的并集组成了 A 的凸壳。注意，我们在不需要进行背景匹配的地方使用了简化的击中或击不中变换，如 9.4 节末尾所论述的那样。

图 9.19 说明了式(9.5.4)和式(9.5.5)给出的过程。图 9.19(a)显示了用于提取凸壳的结构元素。每个元素的原点就是它的中心。“×”项表示“不考虑”条件。意思是，如果在那个位置的结构元素模板下，A 的 3×3 区域匹配了模板模式，则说结构元素找到了 A 中的一个匹配。对一个特定的模板，当 A 中 3×3 区域的中心为 0 时，而在阴影模板元素下的 3 个像素为 1 时，才会出现模式的匹配。不用顾及在 3×3 区域内的其他像素值。同时，图 9.19(a)中的符号 B^i 是将 B^{i-1} 顺时针旋转 90°得到的。

图 9.19(b)显示了一个寻找凸壳的集合 A。从 $X_0^1 = A$ 开始经过式(9.5.4)的 4 次迭代，得到示于图 9.19(c)的集合。然后令 $X_0^2 = A$，并再次使用式(9.5.4)，得到的集合显示在图9.19(d)中(此例子仅经过两步就可以收敛)。

接下来的两个结果以同样的方式得到。最后，图 9.19(c)，(d)，(e)和(f)中集合的并集导出了图 9.19(g)中显示的凸壳。每个结构元素的属性在图 9.19(h)中合成的集合中突显出来。

这个过程一个明显的缺点是，凸壳可能超出确保凸性所需的最小尺寸。减少这种影响的简单方法是，限制生长以便凸壳不会超过初始点集合在水平和垂直方向上的尺寸大小。对图 9.19中的例子使用这种限制的图像结果示于图 9.20 中。更复杂的边界可以进一步限制具有更多细节的图像的生长。例如，我们可以沿着垂直、水平和对角线方向应用原点集的最大尺寸。像这样的改进所付出的代价就是额外增加了算法的复杂性(也增加了对计算量的需求)。

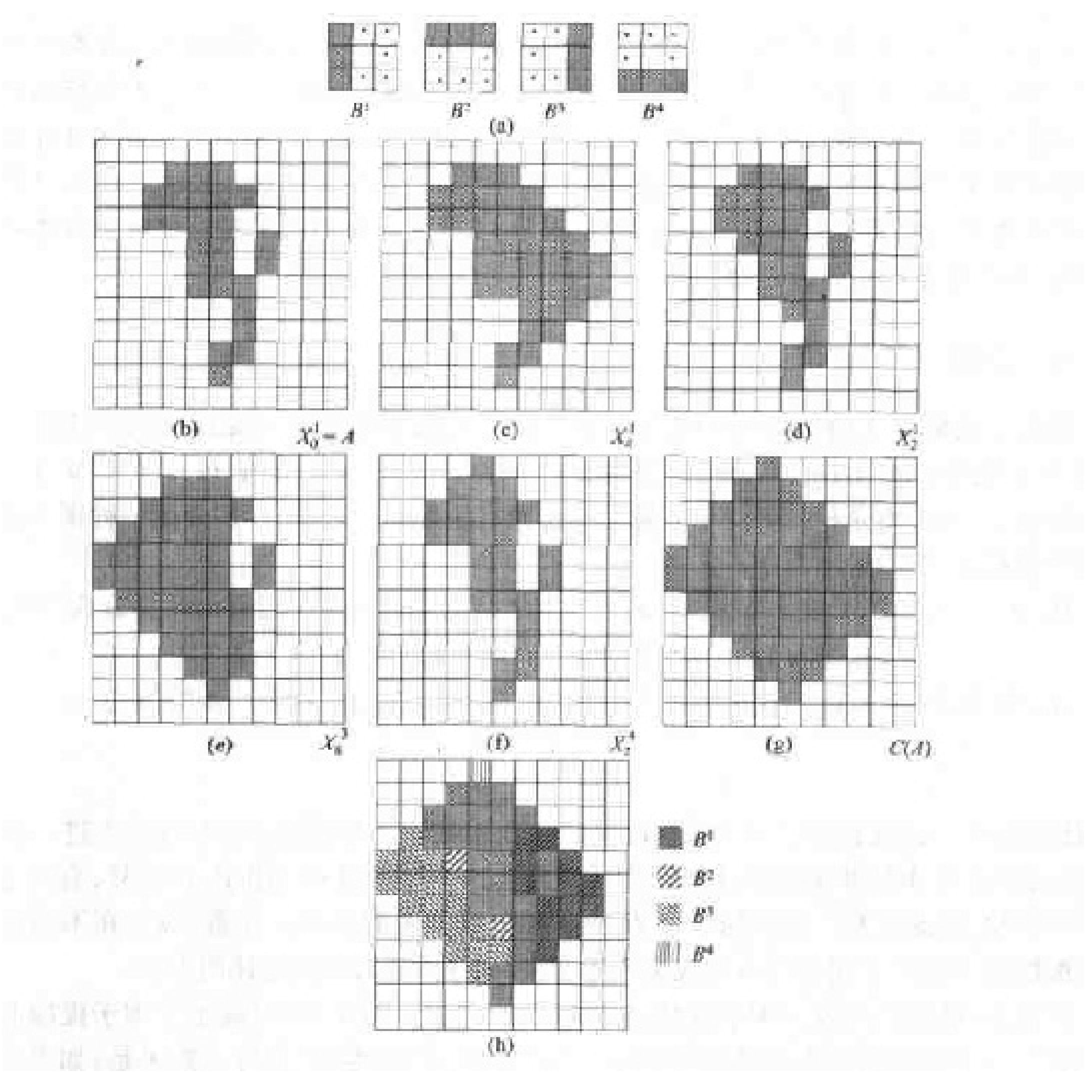


图 9.19 (a)结构元素,(b)集合 A,(c)~(f)用(a)中的结构元素得到的收敛结果,(g)凸壳,(h)凸壳显示了每个结构元素的属性

9.5.5 细化

集合 A 使用结构元素 B 进行细化用 $A\otimes B$ 表示。细化过程可以根据击中或击不中变换定义:

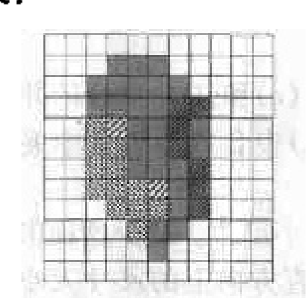


图 9.20 限制凸壳生成,以便不超过初始点集合在水平和垂直方向上的最大尺寸

$$A\otimes B = A-(A \circledast B) = A\cap(A \circledast B)^c \quad (9.5.6)$$

如在前一节那样,我们仅对用结构元素进行模式匹配感兴趣,所以在击中或击不中变换中没有背景运算。相应的对于 A 的细化更为有用的一种表达方式是以结构元素序列为基础的:

$$\{B\} = \{B^1,B^2,B^3,\cdots,B^n\} \quad (9.5.7)$$

这里 B^i 是 B^{i-1}旋转后的形式。使用这个概念,我们现在用结构元素序列定义细化为:

$$A\otimes\{B\} = ((\cdots((A\otimes B^1)\otimes B^2)\cdots)\otimes B^n) \quad (9.5.8)$$

这种处理通过使用 B^1 经一遍处理对 A 进行细化,然后使用

B^2 经一遍处理对得到的结果进行细化，如此进行下去，直到 A 使用 B^n 进行一次细化。整个过程不断重复直到得到的结果不再发生变化。每遍独立的细化过程均使用式(9.5.6)执行。

图 9.21(a)显示了一组通常用于细化的结构元素，图 9.21(b)显示了使用刚才描述的细化过程进行处理的集合 A。图 9.21(c)显示了用 B^1 对 A 进行一遍扫描得到的细化结果。图 9.21(d)到(k)显示了使用其他结构元素处理多遍的结果。在用 B^4 进行第 2 次处理后得到收敛的结果。图 9.21(k)显示了细化的结果。最后图 9.21(l)显示了被转换为 m 连通的细化集合(见 2.5.2 节)以达到消除多重路径的目的。

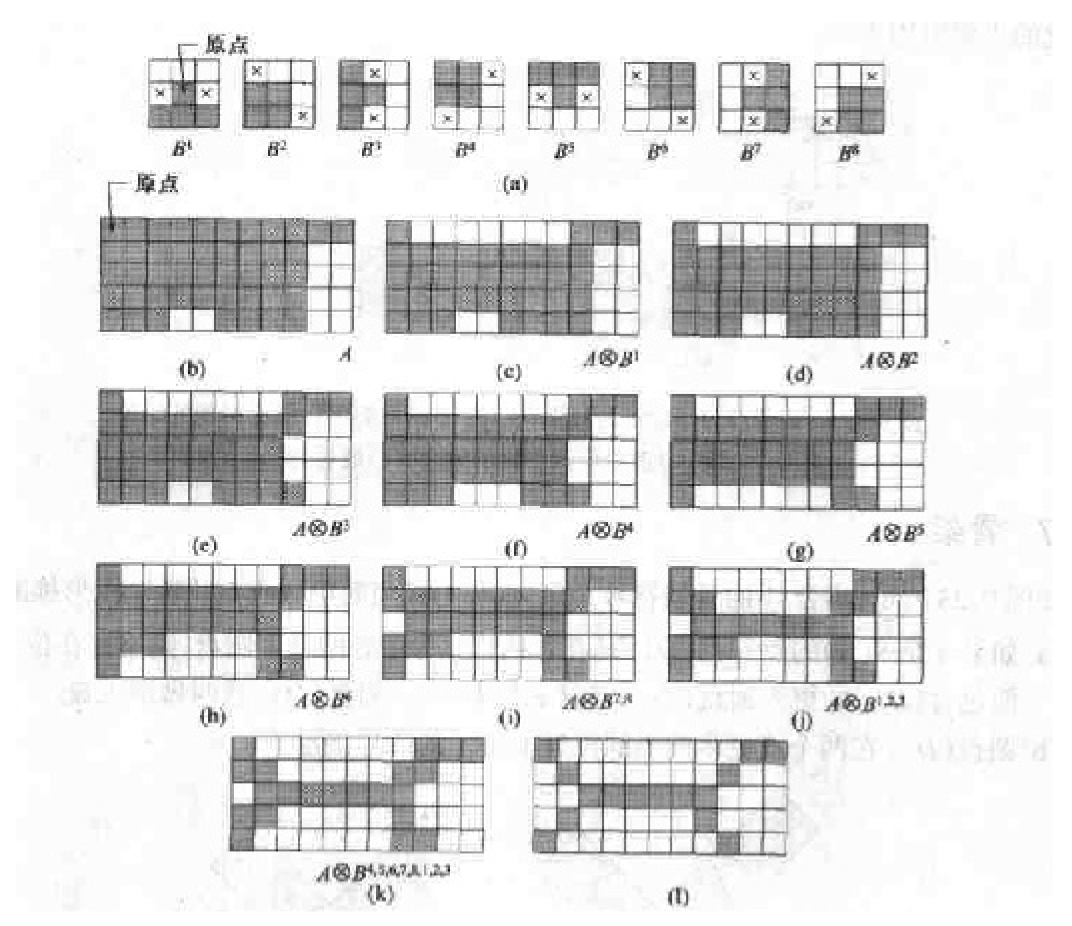


图 9.21　(a)用于细化的经旋转的结构元素序列，(b)集合 A，(c)使用第 1 个结构元素进行细化的结果，(d)～(i)使用接下来的 7 个结构元素进行细化得到的结果(第 7 个和第 8 个结构元素之间没有区别)，(j)再次使用第 1 个结构元素得到的结果(与接下来的两个结构元素没有区别)，(k)收敛后的结果，(l)转换为具有 m 连通度的结果

9.5.6 粗化

粗化与细化在形态学上是对偶过程。它的定义如下：

$$A \odot B = A \cup (A \circledast B) \tag{9.5.9}$$

这里 B 是适合于粗化处理的结构元素。和对细化的定义一样，粗化处理可以定义为一系列操作：

$$A \odot \{B\} = ((\cdots((A \odot B^1) \odot B^2) \cdots) \odot B^n) \tag{9.5.10}$$

如图 9.21(a)所示，用于粗化处理的结构元素和与细化处理有关的结构元素具有相同的形式，

但所有1和0要互换。然而,粗化的分离算法在实际使用中很少用到,代之而来的经常是先对所讨论集合的背景进行细化,而后对结果求补集。换句话说,为了将集合 A 进行粗化,我们先令 $C=A^c$,而后对 C 进行细化,然后再形成 C^c。图9.22对这一过程进行了说明。

根据 A 的性质,这个过程可能产生某些断点,如图9.22(d)所示。所以,使用这种方法的粗化处理通常用简单的后处理消除断点。注意,根据图9.22(c),经细化处理的背景构成了一条边界以备粗化处理。

直接使用式(9.5.10)进行粗化处理不存在这种有用的特性。这是使用背景细化处理来实现粗化的主要原因之一。

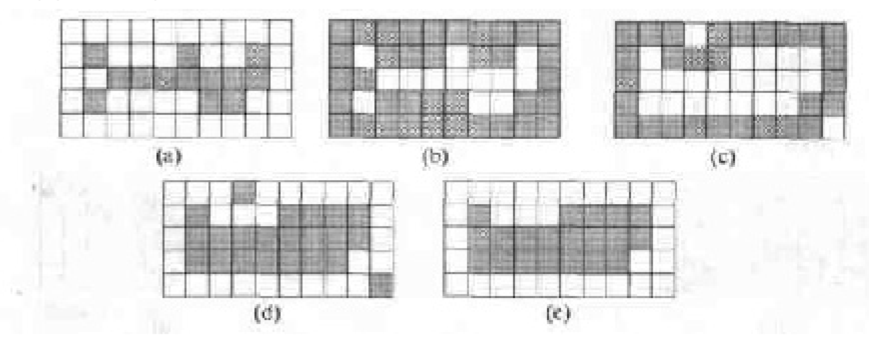


图9.22 (a)集合 A,(b)A 的补集,(c)将 A 的补集进行细化得到的结果,(d)通过对(c)求补得到粗化的集合,(e)最后的结果,没有断点

9.5.7 骨架

如图9.23所示,集合 A 的骨架符号 $S(A)$ 直观上相当简单。我们从这个图形推断:

(a)如果 z 是 $S(A)$ 的点并且 $(D)_z$ 是在 A 内以 z 为中心的最大圆盘,则不存在位于 A 内的能包含 $(D)_z$ 的更大圆盘(不一定以 z 为中心)。圆盘 $(D)_z$ 就叫做最大盘。

(b)圆盘 $(D)_z$ 在两个或更多的不同位置上与 A 的边界接触。

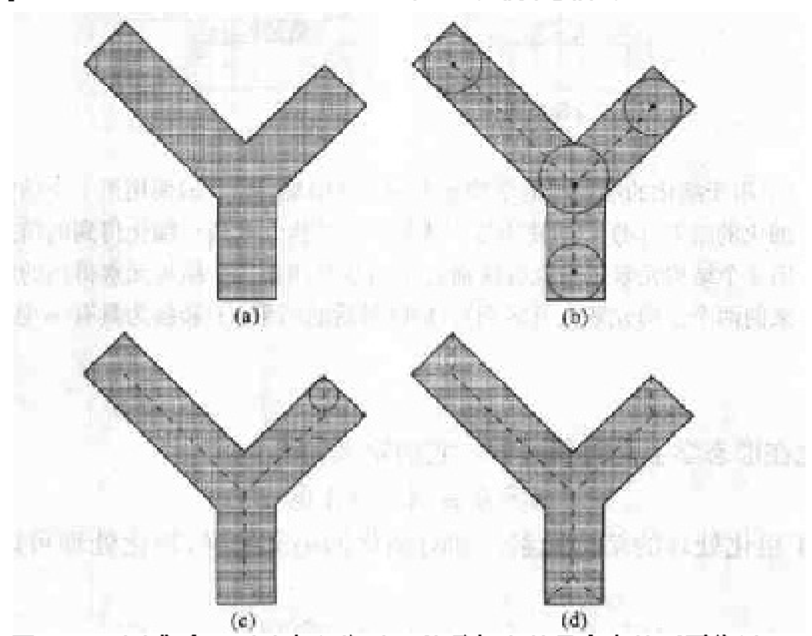


图9.23 (a)集合 A,(b)中心位于 A 的骨架上的最大盘的不同位置,(c)位于 A 的骨架的不同条线段上的最大盘,(d)完整的骨架

A 的骨架可用腐蚀和开操作表达。即,骨架可以表达为如下式所示:

$$S(A) = \bigcup_{k=0}^{K} S_k(A) \tag{9.5.11}$$

而

$$S_k(A) = (A \ominus kB) - (A \ominus kB) \circ B \tag{9.5.12}$$

这里 B 是一个结构元素,$(A \ominus kB)$ 表示对 A 的连续 k 次腐蚀:

$$(A \ominus kB) = (\cdots(A \ominus B) \ominus B) \ominus \cdots) \ominus B \tag{9.5.13}$$

第 k 次是 A 被腐蚀为空集合前进行的最后一次迭代。就是说:

$$K = \max\{k \mid (A \ominus kB) \neq \varnothing\} \tag{9.5.14}$$

式(9.5.11)和式(9.5.12)说明 $S(A)$ 可以由骨架子集 $S_k(A)$ 的并集得到。即,A 可以通过使用下列公式由这些子集重构:

$$A = \bigcup_{k=0}^{K} (S_k(A) \oplus kB) \tag{9.5.15}$$

这里 $(S_k(A) \oplus kB)$ 表示对 $S_k(A)$ 的 k 次连续的膨胀,即,

$$(S_k(A) \oplus kB) = ((\cdots(S_k(A) \oplus B) \oplus B) \oplus \cdots) \oplus B \tag{9.5.16}$$

例 9.8　计算简单图形的骨架

图 9.24 说明了刚才讨论的概念。第 1 列显示了初始集合(在顶部)和使用结构元素 B 的两个腐蚀结果。注意,对 A 再进行一次腐蚀将导致空集出现,因此,此处 $K = 2$。第 2 列显示了使用 B 对第 1 列中的集合进行的开操作。结合图 9.8 中讨论的开操作的拟合特性,这些结果是不难解释的。第 3 列是第 1 列和第 2 列间集合的差。

第 4 列包含两部分骨架和最终结果(在此列的底部)。最终得到的骨架不仅比需要的更粗,而且更重要的是它是不连通的。由于在前面形态学骨架公式化的表达中没有任何条件保证连通性,所以这个结果是我们预料到的。形态学在给定集合的腐蚀和开操作中产生一个精致的公式。然而,如果同通常情况下那样,骨架要求最大限度的细化,需要彼此相连,且受到最小的腐蚀,则就需要像 11.1.5 节所探讨的算法那样的有试探性质的公式。

第 5 列显示了 $S_0(A)$,$S_1(A) \oplus B$ 和 $(S_2(A) \oplus 2B) = (S_2(A) \oplus B) \oplus B$。最终,最后一列显示了重构的集合 A。根据式(9.5.15),这个集合是第 5 列所示骨架子集膨胀后的并集。

9.5.8 裁剪

裁剪方法本质上是对细化处理和骨架绘制算法的补充,因为这些处理过程会将某些附加部分保留下来,而这些部分是应该通过后处理清除干净的。我们从裁剪问题开始讨论并逐渐以前面小节中介绍的材料为基础探讨一种形态学上的解决方法。因此,我们抓住这个机会说明,如何通过到现在为止讨论的几种技术的联合使用解决这一问题。

在自动手写字符识别中通常使用的方法是,分析每种字符的骨架形状。这些骨架经常有各种"毛刺"(寄生成分)。毛刺是在腐蚀过程中由于构成字符的笔画的不均匀性造成的。我们将探讨一种形态学技术来解决这类问题。我们先从假定寄生成分的长度不超过某一确定的像素数开始。

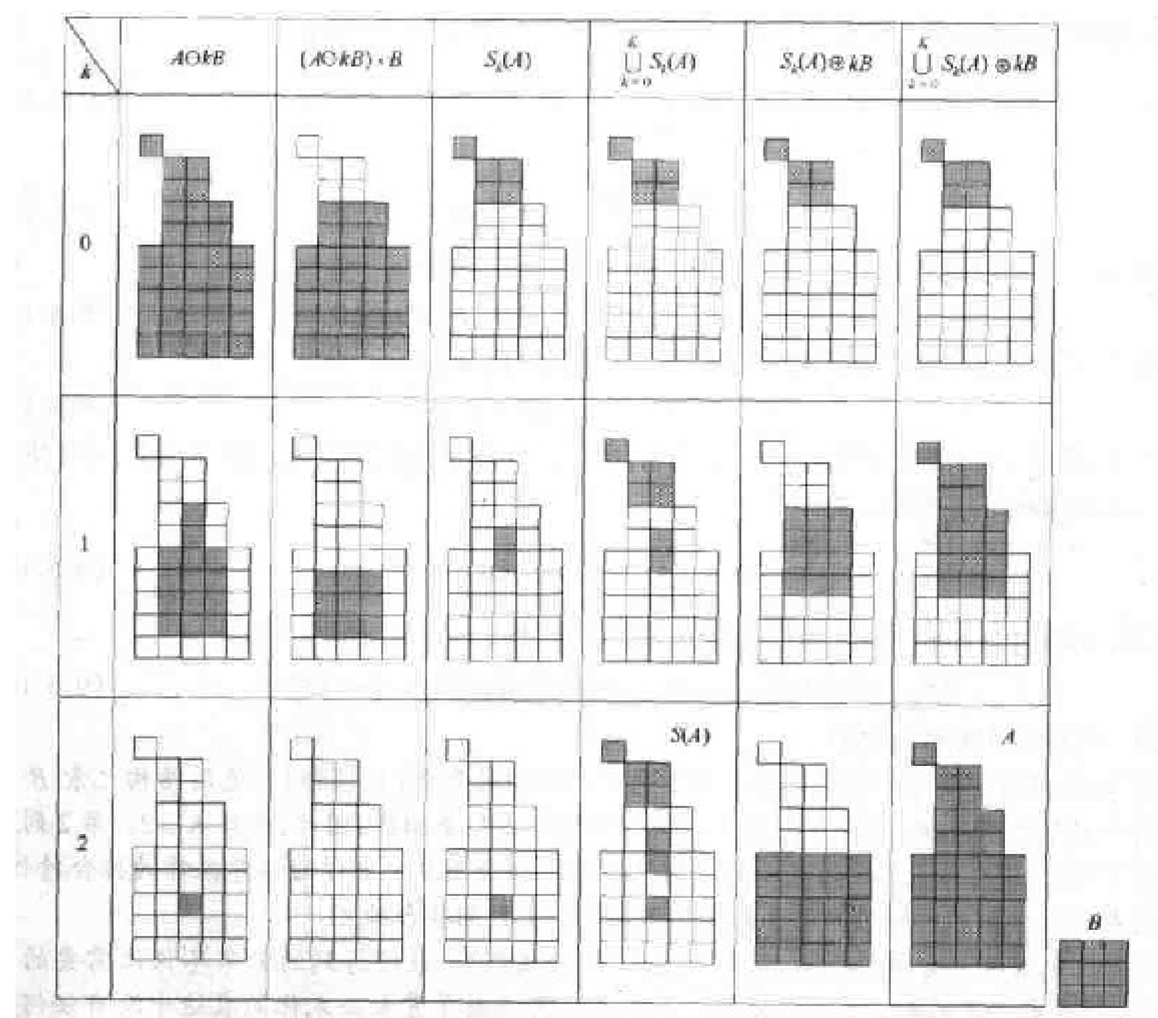


图 9.24 通过式(9.5.15)执行式(9.5.11)。初始集合位于左上角,它的形态学骨架在第4列的底部。第6列底部为重构后的集合

图 9.25(a)显示了一个手写的"a"字符。在字符最左边部分的寄生成分就是我们要消除的部分。解决方案是以不断删除寄生分支的终点为基础的。当然,这会将其字符的其他分支变短,但在缺乏其他结构信息的情况下,我们假定这个例子中任何具有 3 个或小于 3 个像素长度的分支都将被删除。用一系列被设计用来检测终点的结构元素对输人集合 A 进行细化,可以得到希望的结果。即:

$$X_1 = A \otimes \{B\} \tag{9.5.17}$$

这里$\{B\}$表示图 9.25(b)和(c)[见式(9.5.7)有关结构元素序列的部分]中显示的结构元素序列。这个结构元素的序列由两种不同结构组成,每种结构对全部 8 个元素进行 90°的旋转。图 9.25(b)中的"×"表示"不考虑"的情况,在某种意义上,该位置上的像素值是 0 还是 1 并无关系。在有关形态学的著述中的许多报道都是以运用类似图 9.25(b)中的单一结构元素为基础的,只是在第 1 列中多了"不考虑"的条件而已。这是不完善的。例如,这个元素将图 9.25(a)中第 8 行、第 4 列的点识别为终点,因此,将此点消除。这样做就破坏了这个笔画的连续性。

连续对 A 使用 3 次式(9.5.17)得到图 9.25(d)中显示的集合 X_1。下一步是将字符恢复成原来

的形状，但要去掉寄生分支。这首先需要建立在 X_1 中包含的所有终点[图 9.25(e)]的集合 X_2。

$$X_2 = \bigcup_{k=1}^{8} (X_1 \circledast B^k) \tag{9.5.18}$$

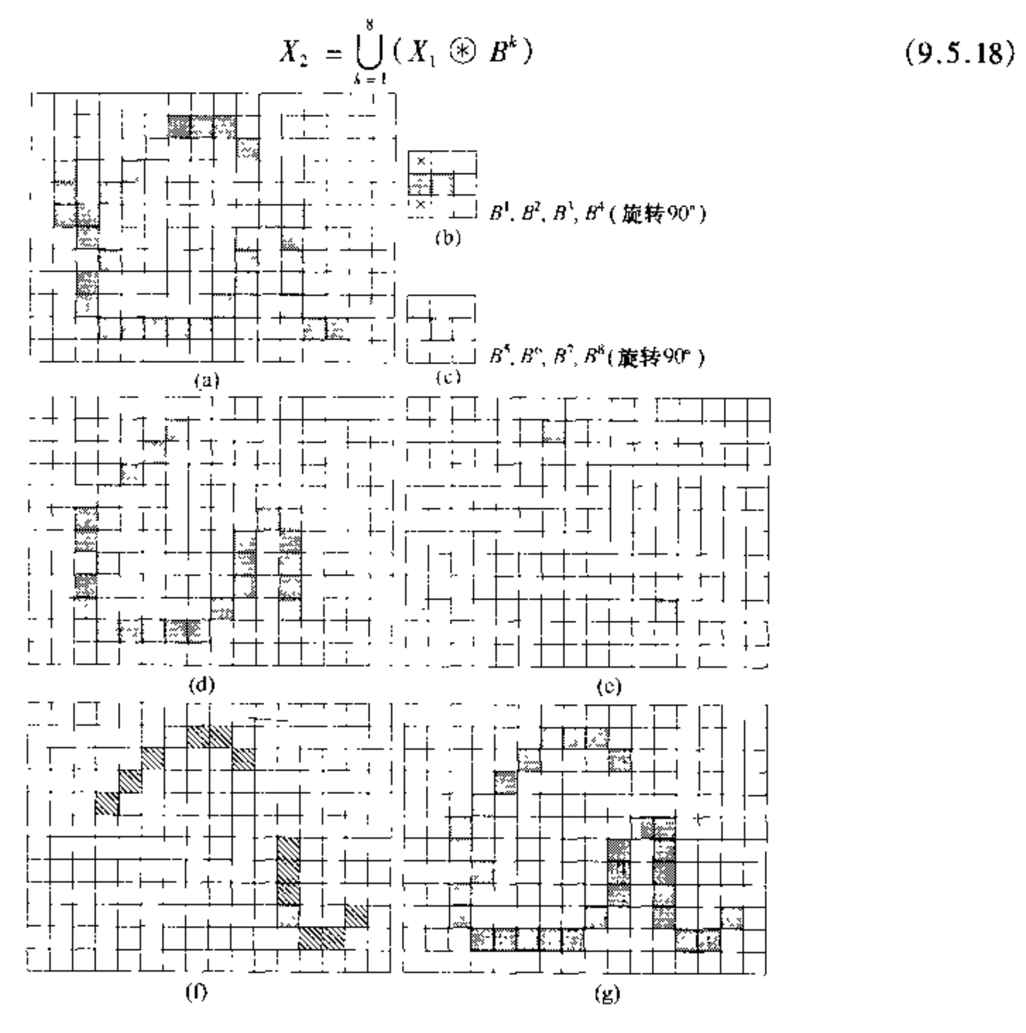


图 9.25　(a)原图，(b)和(c)用于检测终点的结构元素，(d)进行了三轮细化后的结果，(e)(d)的终点，(f)以(a)为条件的终点的膨胀，(g)经过裁剪的图像

这里 B^k 是示于图 9.25(b)和(c)中相同的端点检测子。下一步是对端点进行三次膨胀处理，用集合 A 作为消减因子：

$$X_3 = (X_2 \oplus H) \cap A \tag{9.5.19}$$

这里，H 是元素值为 1 的 3×3 的结构元素。在这种区域填充和提取连通分量的情况下，条件膨胀处理可以防止在我们关注的区域外产生值为 1 的元素。图 9.25(f)中显示的结果可以作为证明。最后，X_3 和 X_1 的并集：

$$X_4 = X_1 \cup X_3 \tag{9.5.20}$$

得到我们想要的结果，如图 9.25(g)所示。

在更复杂的情况下，使用式(9.5.19)有时会得到某些寄生分支的"尖端"。当这些分支的端点离骨架很近时，这种情况就会发生。尽管使用式(9.5.17)可以消除它们，但由于它们是 A 中的有效点，所以在膨胀过程中会再次捡回这些点（如果这些寄生元素相对于有效的笔画较短，这种情况是很少发生的）。如果全部寄生分支再次被捡回，检测并消除这些点是很容易的，因为它们都是不连续区域。

此处,一种必然的想法就是一定存在一种更为简单的方法来解决这个问题。例如,我们可以留意所有被删除的点并将其中合适的点与那些应用式(9.5.17)后留下的端点进行再连接。这样的选择是正确的。前述这种方式的优点就是,使用简单的形态结构可以解决整个问题。在实际情况下,当这一系列工具有效时,这种选择的好处就是不必写新的算法。我们只需将必要的形态函数结合到运算序列中就可以了。

9.5.9 关于二值图像的形态学运算总结

表9.2总结了前面章节中探讨的形态学上的成果。图9.26总结了迄今为止用于不同形态学处理的结构元素的基本类型。注意,表9.2的第3列中的罗马数字是指图9.26中的结构元素。

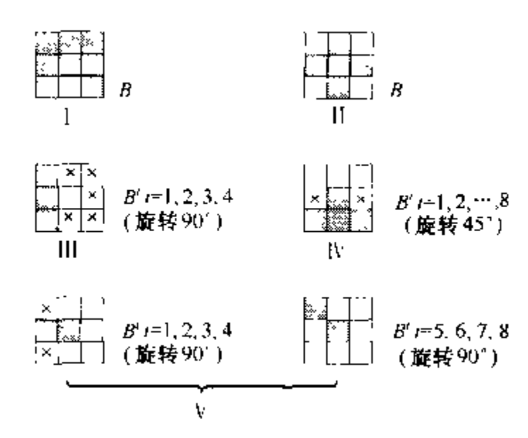


图9.26 用于二值形态学的5种基本结构元素。每种元素的原点是它本身的中心。×表示不考虑的元素

表9.2 形态学操作及其性质的总结

操作类型	等式	评述(罗马数字是指图9.26中的结构元素)
平移	$(A)_z = \{w \mid w = a + z, a \in A\}$	将 A 的原点平移到 z 点
映射	$\hat{B} = \{w \mid w = -b, b \in B\}$	将所有 B 的元素进行相对于 B 的原点的映射
求补集	$A^c = \{w \mid w \notin A\}$	不属于 A 的点的集合
求差集	$A - B = \{w \mid w \in A, w \notin B\} = A \cap B^c$	属于 A 但不属于 B 的点的集合
膨胀	$A \oplus B = \{z \mid (\hat{B})_z \cap A \neq \varnothing\}$	"扩展" A 的边界(Ⅰ)
腐蚀	$A \ominus B = \{z \mid (B)_z \subseteq A\}$	"收缩" A 的边界(Ⅰ)
开操作	$A \circ B = (A \ominus B) \oplus B$	平滑轮廓,切断狭区,消除小的"孤岛"和突刺(Ⅰ)
闭操作	$A \cdot B = (A \oplus B) \ominus B$	平滑轮廓,融合狭窄的间断和细长的"沟壑",消除小的孔洞(Ⅰ)
击中或击不中变换	$A \circledast B = (A \ominus B_1) \cap (A^c \ominus B_2)$ $= (A \ominus B_1) - (A \oplus \hat{B}_2)$	点(坐标)的集合。在集合中的点的位置上能同时得到 B_1 在 A 中的一个匹配(击中)以及 B_2 在 A^c 中的一个匹配
边界提取	$\beta(A) = A - (A \ominus B)$	在集合 A 边界上的点的集合(Ⅰ)
区域填充	$X_k = (X_{k-1} \oplus B) \cap A^c; X_0 = p, k = 1,2,3,\cdots$	给定 A 中某个区域内的一个点 p,填充此区域(Ⅱ)

（续表）

操作类型	等　式	评述(罗马数字是指图9.26中的结构元素)
连通分量	$X_k = (X_{k-1} \oplus B) \cap A$; $X_0 = p$, $k = 1,2,3,\cdots$	给定 Y 内的一个点 p，寻找 A 中的连通分量 Y(Ⅰ)
凸壳	$X_k^i = (X_{k-1}^i \circledast B^i) \cup A$; $i = 1,2,3,4$; $k = 1,2,3,\cdots$; $X_0^i = A$; $D^i = X_{conv}^i$	寻找集合 A 的凸壳 $C(A)$，这里"conv"表示在 $X_k^i = X_{k-1}^i$ 的意义上收敛(Ⅲ)
细化	$A \otimes B = A - (A \circledast B) = A \cap (A \circledast B)^c$ $A \otimes \{B\} = ((\cdots((A \otimes B^1) \otimes B^2)\cdots) \otimes B^n)$ $\{B\} = \{B^1, B^2, B^3, \cdots, B^n\}$	细化集合 A。前两个等式给出细化的基本定义，后两个等式表示用结构元素序列进行细化。在实际应用中通常使用这种方法(Ⅳ)
粗化	$A \odot B = A \cup (A \circledast B)$ $A \odot \{B\} = ((\cdots(A \odot B^1) \odot B^2 \cdots) \odot B^n)$	粗化集合 A(见前面关于结构元素序列的评述)，使用Ⅳ，0和1位置互换即可
骨架	$S(A) = \bigcup_{k=0}^{K} S_k(A)$ $S_k(A) = \bigcup_{k=0}^{K} \{(A \ominus kB) - [(A \ominus kB) \circ B]\}$ A 的重构： $A = \bigcup_{k=0}^{K}(S_k(A) \oplus kB)$	寻找集合 $S(A)$ 的骨架，最后的等式表示 A 可以由它的骨架子集进行重构。在这3个等式中，K 是迭代步骤的次数。超过 K 次迭代，集合 A 将被腐蚀为空集，符号 $(A \ominus kB)$ 表示对 A 用 B 的第 k 次迭代，进行连续腐蚀(Ⅰ)
裁剪	$X_1 = A \otimes \{B\}$ $X_2 = \bigcup_{k=1}^{8}(X_1 \circledast B^k)$ $X_3 = (X_2 \oplus H) \cap A$ $X_4 = X_1 \cup X_3$	X_4 是裁剪集合 A 的结果。用第1个等式求取 X_1 所使用的次数必须是确定的。结构元素Ⅴ用于前两个等式。第3个等式中，H 表示结构元素Ⅰ

9.6　灰度级图像扩展

在本节，我们把形态学处理扩展到灰度图像的基本操作，即膨胀、腐蚀、开操作和闭操作。然后，我们利用这些操作探讨几种基本的灰度级形态学算法。特别是要建立通过形态学梯度运算进行边界提取的算法，以及基于纹理内容的区域分割算法。同时，也将讨论平滑处理和尖锐化处理的算法，这些算法在预处理和后处理过程中非常有用。

通过下边的讨论，我们将处理形如 $f(x,y)$ 和 $b(x,y)$ 的数字图像函数，这里 $f(x,y)$ 是输入图像，而 $b(x,y)$ 是结构元素，$b(x,y)$ 本身是一个子图像函数。假设这些函数是在2.4.2节介绍过的离散函数。即，如果 Z 表示实整数集合，则假设 (x,y) 是来自 $Z \times Z$ 的整数，且 f 和 b 是对每一个 (x,y) 坐标赋以灰度值(来自实数集合 R 的实数)的函数。如果灰度级也是整数，则用 Z 代替 R。

9.6.1 膨胀

用 b 对函数 f 进行的灰度膨胀表示为:$f \oplus b$,定义为:

$$(f \oplus b)(s,t) = \max\{f(s-x,t-y)+b(x,y) \mid (s-x),(t-y) \in D_f;(x,y) \in D_b\} \tag{9.6.1}$$

这里 D_f 和 D_b 分别是 f 和 b 的定义域。请记住,f 和 b 是函数而不是二值形态学情况中的集合。

$(s-x)$和$(t-y)$必须在 f 的定义域内以及 x 和 y 必须在 b 的定义域内的条件与膨胀的二值定义中的条件是相似的(这里两个集合的交集至少应有一个元素)。同时应该注意式(9.6.1)的形式与二维卷积[式(4.2.30)]是相似的,并且用最大值运算代替卷积求和,用加法运算代替卷积乘积。

下面我们将用简单的一维函数说明式(9.6.1)的表示法和运算原理。对单变量函数,式(9.6.1)简化为表达式:

$$(f \oplus b)(s) = \max\{f(s-x)+b(x) \mid (s-x) \in D_f, x \in D_b\}$$

回顾关于卷积的讨论:$f(-x)$是 $f(x)$关于 x 轴原点的镜像。在卷积运算中,s 为正时函数 $f(s-x)$向右移动,s 为负则向左移动。条件是$(s-x)$必须在 f 的定义域内,x 的值必须在 b 的定义域内,这意味着 f 和 b 是彼此交叠的。正如在上一段中提到的,这些条件与二值图像膨胀的条件(这两个集合的交集至少有一个元素)是相似的。最后,与二值图像的情况不同,被移动的是 f 而不是 b。式(9.6.1)可以被写成 b 受到平移而不是f。然而,如果 D_b 比 D_f 小(实际应用中经常是这样的),则式(9.6.1)中给出形式的索引项可以进一步简化而取得相同的结果。从概念上讲,以 b 滑过函数f还是以f滑过 b 是没有区别的。事实上,尽管这个等式更容易实现,但如果以 b 作为滑过f的函数,则对灰度膨胀的实际原理在直观上更容易理解一些。

图 9.27 中显示了一个例子。注意,在每个结构元素的位置上,这一点的膨胀值是在跨度为 b 的区间内f与 b 之和的最大值。通常对灰度图像进行膨胀处理的结果是双重的:(1)如果所有结构元素的值为正,则输出图像会趋向于比输入图像更亮;(2)暗的细节部分全部减少了还是被消除掉了取决于膨胀所用的结构元素的值和形状。

9.6.2 腐蚀

灰度腐蚀,表示为 $f \ominus b$,定义为:

$$(f \ominus b)(s,t) = \min\{f(s+x,t+y)-b(x,y) \mid (s+x),(t+y) \in D_f;(x,y) \in D_b\} \tag{9.6.2}$$

这里 D_f 和D_b 分别是f 和b 的定义域。平移参数$(s+x)$和$(t+y)$必须在f的定义域内,而且 x 和 y 必须在 b 的定义域内,这与腐蚀的二值定义中的条件(这里结构元素必须完全包含在被腐蚀的集合内)相似。注意,式(9.6.2)在形式上与二维相关[式(4.6.30)]是相似的,并且用最小值运算代替了相关运算,用减法运算代替了相关的乘积。

我们通过对一个简单的一维函数进行腐蚀来说明式(9.6.2)的原理。对单变量函数,腐蚀的表达式简化为:

$$(f \ominus b)(s) = \min\{f(s+x)-b(x) \mid (s+x) \in D_f, x \in D_b\}$$

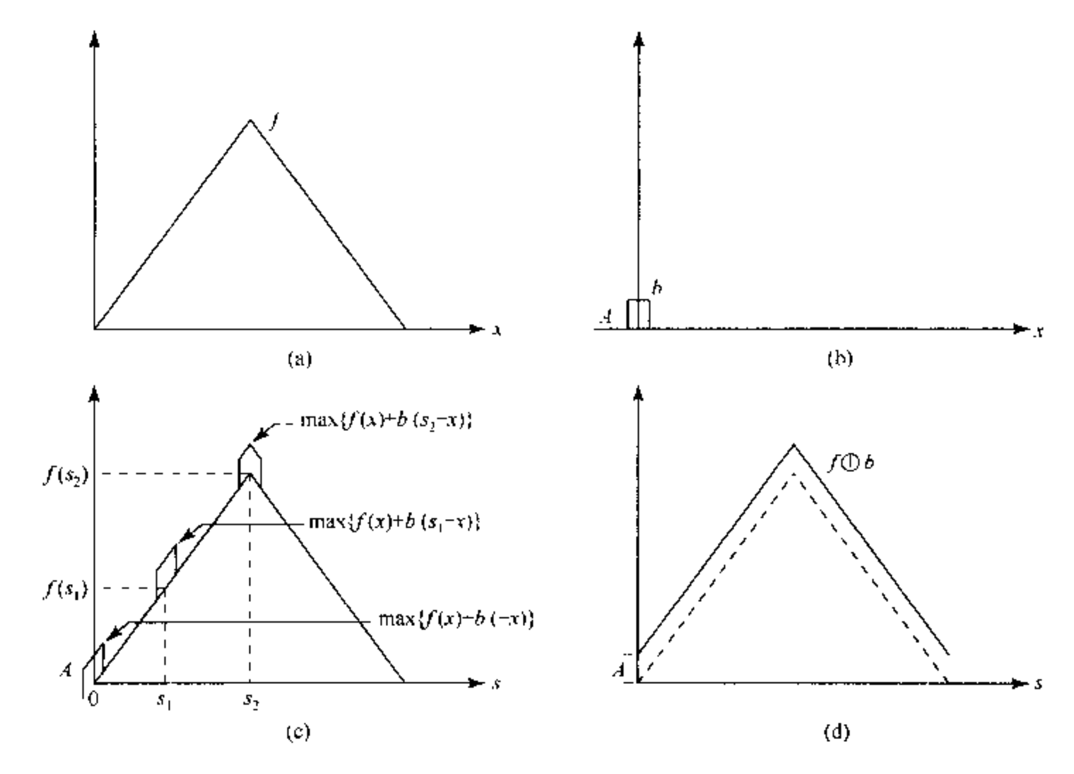


图 9.27 (a)一个简单的函数,(b)高度 A 的结构元素,(c)b 滑过 f 的不同位置进行膨胀的结果,(d)膨胀得到的完整结果(以实线表示)

与相关运算一样,对正的 s,函数 $f(s+x)$ 向左移动,对负的 s,函数向右移动。对 $(s+x)\in D_f$ 和 $x\in D_b$ 的要求表明 b 的取值范围完全包含在移动后的 f 的范围之内。正如前一段提到的,这些要求与腐蚀的二值定义中的条件(这里结构元素必须完全包含在被腐蚀的集合之内)相似。

最后,与腐蚀的二值定义不同,移动的对象是 f 而不是结构元素 b。式(9.6.2)可以写成 b 是被平移的,但会导致表达式的下标索引变得更为复杂。因为 f 在 b 上滑动和 b 在 f 上滑动概念上是相同的,所以,由于在膨胀讨论的结尾处提到的原因使用了式(9.6.2)现在的形式。图 9.28 显示了使用图 9.27(b)的结构元素对图 9.27(a)的函数进行腐蚀的结果。

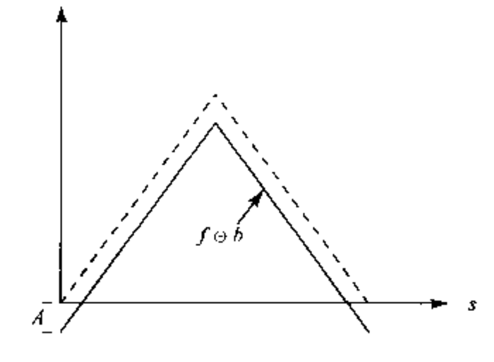


图 9.28 使用图 9.27(b)中显示的结构元素对图 9.27(a)中显示的函数进行腐蚀

式(9.6.2)说明了腐蚀操作是以在结构元素形状定义的区间中选取 $(f-b)$ 最小值为基础的。通常对灰度图像进行腐蚀是双重的:(1)如果所有的结构元素都为正,输出图像会趋向于比输入图像更暗;(2)在输入图像中亮的细节的面积如果比结构元素的面积小,则亮的效果将被消弱。消弱的程度取决于环绕于亮细节周围的灰度值和结构元素自身的形状与幅值。

与函数求补和映射相关的灰度膨胀和腐蚀是对偶的,即:

$$(f\ominus b)^c(s,t)=(f^c\oplus\hat{b})(s,t) \tag{9.6.3}$$

这里 $f^c=-f(x,y)$,$\hat{b}=b(-x,-y)$。除非需要声明,我们在下列讨论中将忽略所有函数的变量以简化表示法。

例 9.9 灰度图像的膨胀和腐蚀的说明

图 9.29(a)显示了一幅 512×512 大小的灰度图像,图 9.29(b)显示了用"平顶"结构元素对图像进行膨胀的结果,使用的结构元素为单位高度且大小为 5×5 像素的平行六面体。基于前面的讨论,我们希望通过膨胀得到的图像比原图像更明亮并且减弱或消除小的、暗的细节部分。

这些效果在图 9.29(b)中清晰可见。不仅得到的图像比原图像更明亮,而且黑色部分已经被减弱了。比如,马的鼻孔部分、从马耳垂到马颈用纽饰装饰的缰绳的黑色部分都有明显的减弱。图 9.29(c)显示了对原图进行腐蚀的结果。注意与膨胀相反的效果。被腐蚀的图像更暗,并且尺寸小、明亮的部分(比如缰绳上的纽饰)被削弱了。

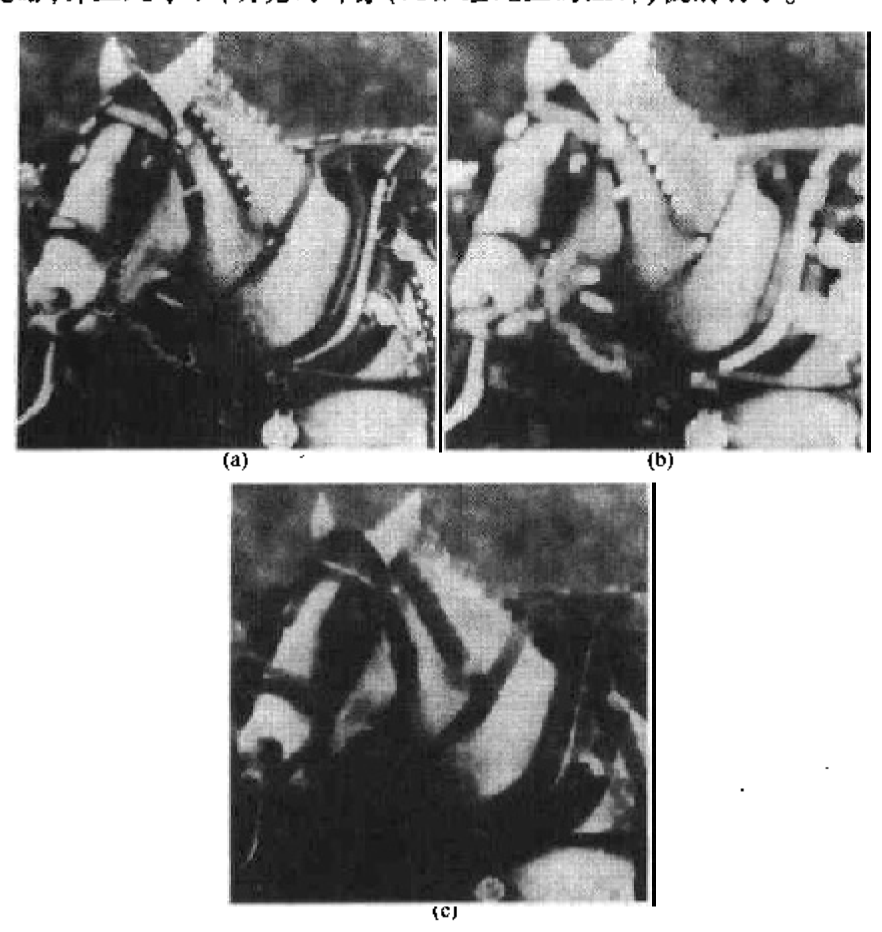


图 9.29 (a)原图,(b)膨胀后的结果,(c)腐蚀后的结果
(由 Leica Cambridge 公司的 A. Morris 先生提供)

9.6.3 开操作和闭操作

灰度图像的开操作和闭操作与二值图像的对应操作具有相同的形式。用子图(结构元素) b 对图像 f 进行开操作表示为 $f\circ b$,定义为:

$$f\circ b=(f\ominus b)\oplus b \tag{9.6.4}$$

同二值图像中的情况一样,开操作先用 b 对 f 进行简单的腐蚀操作,而后用 b 对得到的结果进行膨胀操作。同样,用 b 对 f 进行闭操作表示为 $f\cdot b$,定义为:

$$f \cdot b = (f \oplus b) \ominus b \tag{9.6.5}$$

灰度图像的开操作和闭操作对于求补和映射运算是对偶的。即：

$$(f \cdot b)^c = f^c \circ \hat{b} \tag{9.6.6}$$

因为 $f^c = -f(x,y)$，式(9.6.6)可以写成 $-(f \cdot b) = -(f \circ \hat{b})$ 的形式。

图像的开操作和闭操作具有简单的几何解释，假设在三维透视空间中我们观察一个图像函数 $f(x,y)$（比如地形图），x 轴、y 轴是通常意义上的空间坐标，第 3 个轴是灰度值。在这个坐标系中，图像呈现不连续曲面的形态，曲面上任意点 (x,y) 的值是这个坐标上 f 的值。假设我们使用球形结构元素 b 对 f 进行开操作，将这个结构元素视为"滚动的"球。用 b 对 f 进行开操作的原理可以在几何上解释为，推动球沿着曲面的下侧面滚动，以便球体能在曲面的整个下侧面来回移动。当球体滚过 f 的整个下侧面时，由球体的任何部分接触到的曲面的最高点就构成了开操作 $f \circ b$ 的曲面。

图 9.30 对这个概念进行了说明。为了对说明进行简化，图 9.30(a)将灰度图像的扫描线显示成一个连续函数。图 9.30(b)显示了不同位置上滚动的球。图 9.30(c)显示了沿着扫描线用 b 对 f 进行开操作得到的完整结果。所有比球体直径窄的波峰在幅度和尖锐程度上都减小了。实际应用中，开操作经常用于去除较小（相对于结构元素的大小而言）的明亮细节，同时相对地保持整体的灰度级和较大的明亮区域不变。先进行腐蚀操作可以除去小的图像细节，但这样做会使图像变暗。接下来进行膨胀操作又会增强图像的整体亮度，但不会将腐蚀操作除去的部分重新引进图像中去。

图 9.30(d)和(e)显示了用 b 对 f 进行闭操作的结果。这里，球体在曲面的上表面滑动，波峰基本上保持了它们原来的形状（假设在波峰间最窄的地方超过了球体的直径）。实际上，闭操作经常用于除去图像中的暗细节部分，而相对地保持明亮部分不受影响。先通过膨胀除去图像中的暗细节，同时增加图像的亮度，接下来对暗图像进行腐蚀，而不会将膨胀操作除去的部分重新引入图像之中。将图 9.30 和图 9.8、图 9.9 进行对比是很有意思的。

灰度级开操作满足下列性质：

(i) $(f \circ b) \dashv f$

(ii) 如果 $f_1 \dashv f_2$，则 $(f_1 \circ b) \dashv (f_2 \circ b)$

(iii) $(f \circ b) \circ b = f \circ b$

符号 $e \dashv r$ 表示 e 的域是 r 的域的子集，且对 e 的域内的任何 (x,y) 有 $e(x,y) \leq r(x,y)$。

同样，闭操作满足下列性质：

(i) $f \dashv (f \cdot b)$

(ii) 如果 $f_1 \dashv f_2$，则 $(f_1 \cdot b) \dashv (f_2 \cdot b)$

(iii) $(f \cdot b) \cdot b = f \cdot b$

这些表达式的效用同二值图像对应的性质相似。

例 9.10　灰度级的开操作和闭操作的说明

图 9.31(a)显示了使用与图 9.29(a)中相同的结构元素对图像 9.29(a)进行开操作的结果。注意，小的、明亮的细节尺寸变小了，而暗的、灰度的效果没有明显的变化。图9.31(b)显

示了对图9.29(a)进行闭操作的结果。注意,小的暗细节的尺寸被缩小了,而相对来说,明亮的部分受到的影响很小。

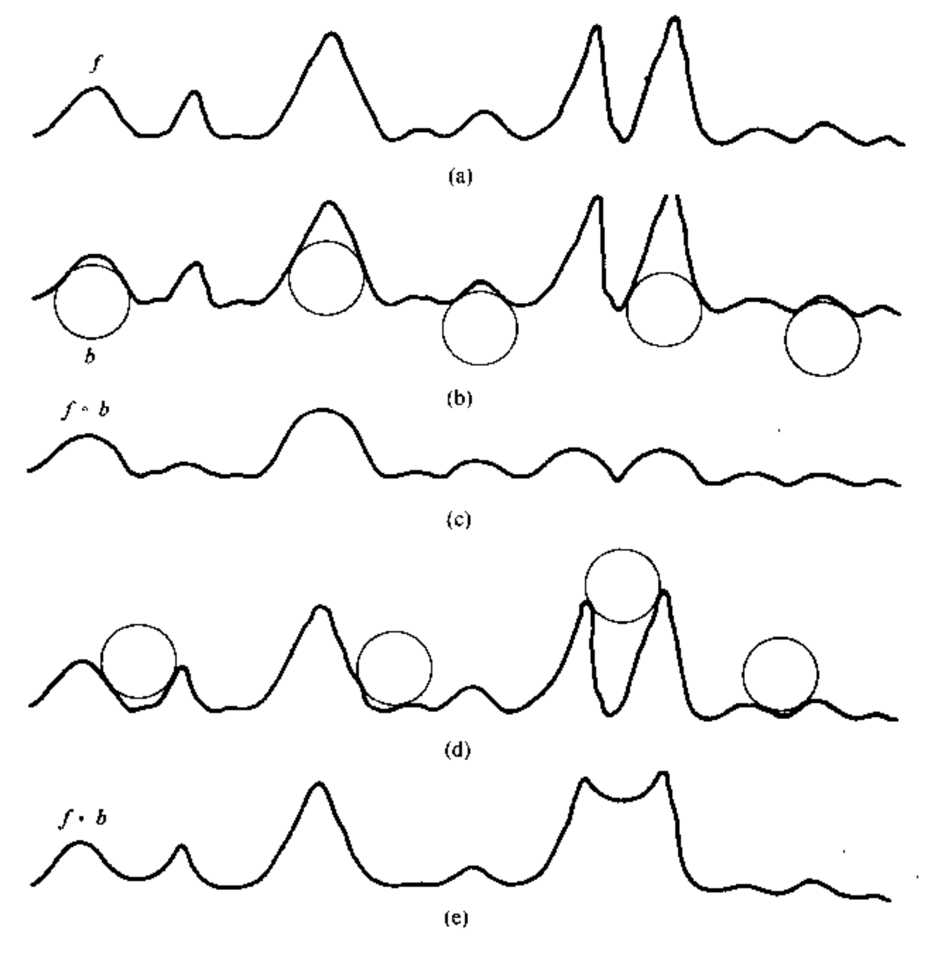


图 9.30 (a) 一条灰度扫描线,(b)开操作时滚动球的位置,(c)开操作的结果,(d)闭操作时滚动球的位置,(e)闭操作的结果

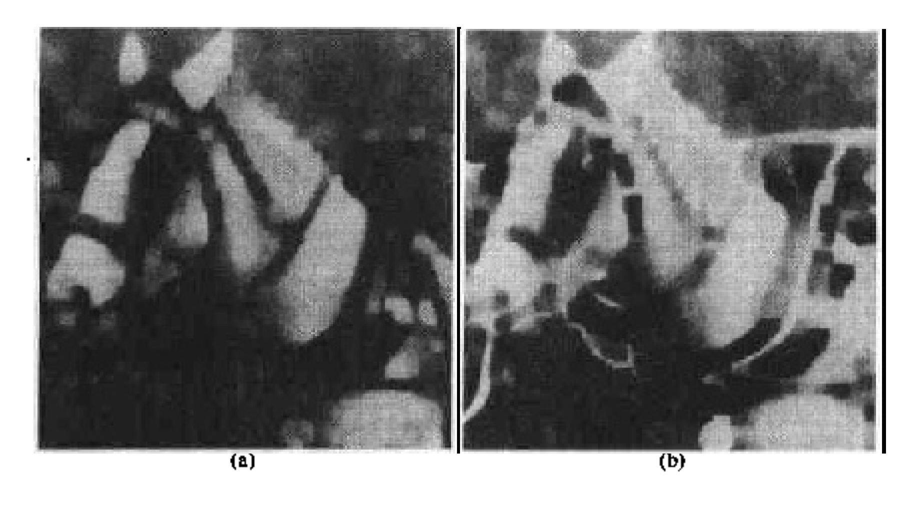


图 9.31 (a)图 9.29(a)的开操作,(b)图 9.29(a)的闭操作
(由 Leica Cambridge 公司的 A. Morris 先生提供)

9.6.4 灰度级形态学的一些应用

我们通过介绍灰度级形态学各种应用的某些细节对有关形态学技术进行总结。除非另行声明，所有图像均为 512×512 像素大小，并且经过与图 9.29 有关的结构元素的处理。

形态学图像平滑处理

实现平滑处理的一种途径是先使用形态学开操作，而后进行闭操作。进行这两种操作的最终结果是除去或减少了人为亮和暗的因素或噪声。图 9.32 是对示于图 9.29(a)的图像进行平滑处理后得到的图像。



图 9.32　对图 9.29(a)中的图像进行形态学平滑（由 Leica Cambridge 公司的 A. Morris 先生提供）

形态学图像梯度

除了前面讨论的消除由于人为因素产生的小的黑和亮区域外，膨胀和腐蚀经常用于计算图像的形态学梯度，表示为 g：

$$g = (f \oplus b) - (f \ominus b) \tag{9.6.7}$$

图 9.33 显示了计算图 9.29(a)中图像的形态学梯度得到的结果。正如我们希望的，形态学梯度使输入图像中灰度级的跃变更为急剧。与用 3.7.3 节中讨论的方法求取的梯度相反，使用对称结构元素得到的形态学梯度对边缘方向性的依赖更小。

top-hat 变换

对图像进行所谓的形态学 top-hat 变换，用 h 表示，定义为：

$$h = f - (f \circ b) \tag{9.6.8}$$

如以前一样，f 是输入图像，b 是结构元素函数。最初对这种变换命名是由于使用的带有一个平顶的圆柱形或平行六面体形的结构元素。它对于增强阴影的细节很有用处。图 9.34 显示了对图 9.29(a)中图像进行 top-hat 变换得到的结果。注意，马头下部背景区域的细节得到了增强。

纹理分割

图 9.35(a)显示了一幅由两个纹理区域组成的简单灰度图像。右边区域包含的圆点的直径比左边圆点的大。目的是以纹理为基础找到两个区域的边界。

因为闭操作趋向于消除图像中的暗细节，所以在这种特定的情况下，处理过程应依次从小到大使用不同大小的结构元素对输入图像进行闭操作。当结构元素的尺寸与小斑点的大小相当时，可以将这些斑点从图像中除去，在这些斑点的位置上只留下浅色的背景。

在处理的这一步上，只剩下比结构元素更大的斑点和左边大斑点之间的浅色背景。下一步，使用比大斑点之间的间隙大的结构元素进行一次开操作。这个过程除去斑点之间浅色的色块，而留下右边包含大的暗斑点及斑点间相同的暗色块的暗区域。在这一步，处理过程在左边生成了浅色区域而在右边生成暗区域。使用一个简单门限就可以得到两个纹理区域之间的

边界。图 9.35(b)显示了叠加在原图上的边界结果。使用图 9.30 中描述的滚动球进一步在细节上完成这个例子是很有启发性的。

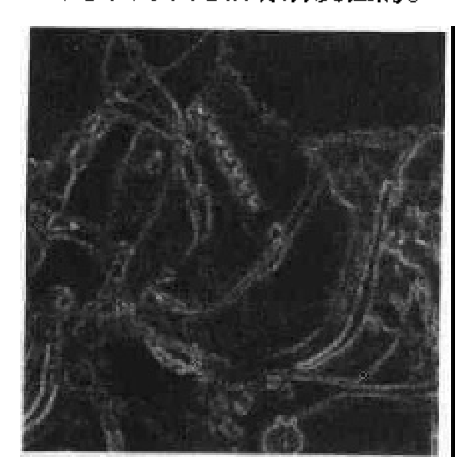


图 9.33 图 9.29(a)中图像的形态学梯度(由 Leica Cambridge 公司的 A. Morris 先生提供)



图 9.34 对图 9.29(a)中的图像进行一次 top-hat 变换得到的结果(由 Leica Cambridge 公司的 A. Morris 先生提供)

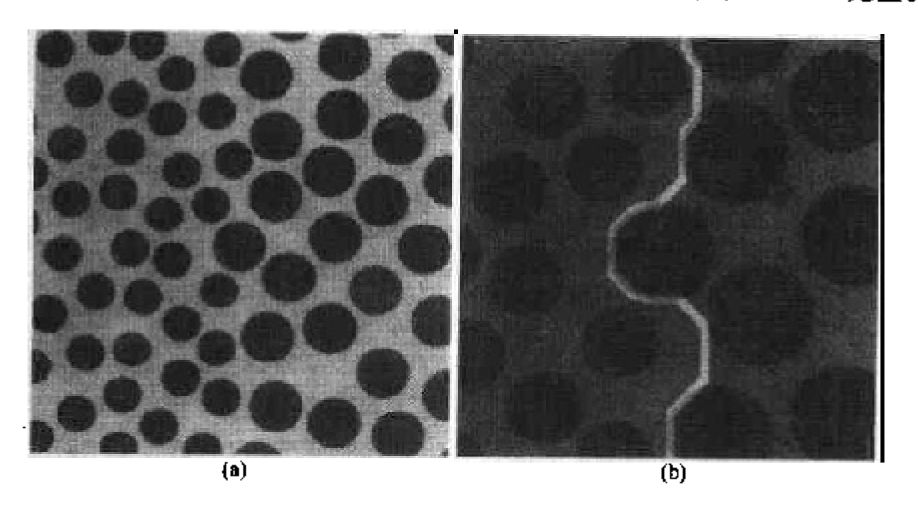
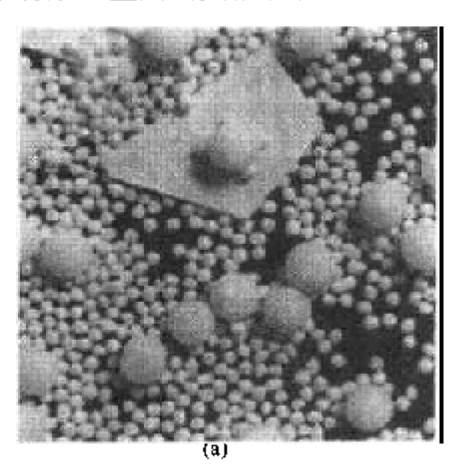


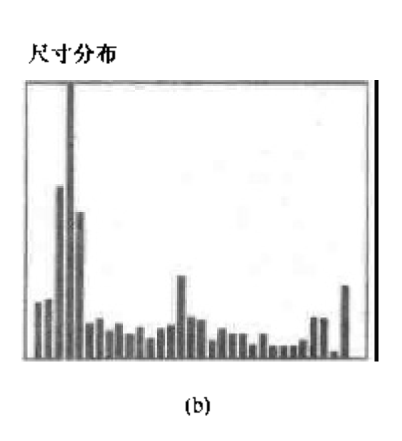
图 9.35 (a)原图像,(b)显示不同纹理区域间边界的图像(由 Leica Cambridge 公司的 A. Morris 先生提供)

粒度测定

粒度测定所处理的主要领域是判断图像中颗粒的尺寸分布问题。图 9.36(a)显示了三种不同尺寸亮目标组成的图像。这些目标不仅相互交叠,而且排列散乱难于识别单个颗粒。因为颗粒比背景稍微亮一些,下面的形态学方法可以用于检测尺寸分布。使用逐渐增大尺寸的结构元素对原图进行开操作。当每一次使用不同尺寸的结构元素处理之后,初始图像和经过开操作处理的图像之间的差异可以计算出来。在处理的最后阶段,将这些计算的差异进行归一化处理,然后建立颗粒尺寸分布的直方图。这种处理方法是基于下面的思想:以某一特定尺寸对含有相近尺寸颗粒的图像区域进行开操作,对输入图像得到的处理效果最好。因此,通过计算输入和输出图像之间的差异可以对相近尺寸颗粒的相对数量进行测算。图9.36(b)显示

了在这种情况得到的尺寸分布。直方图表明在输入图像中存在三种主要尺寸的颗粒。这种处理对于检测带有某一主要的类似颗粒状特征的区域是很有用处的。


(a)



(b)

图 9.36　(a)包含彼此重叠的颗粒的原图像,(b)尺寸分布
(由 Leica Cambridge 公司的 A. Morris 先生提供)

小结

本章中介绍的形态学概念和技术构成了一系列能从图像中提取某种感兴趣特征的有力工具集。形态学图像处理最吸引人的一个方面就是根据形态学技术已有的发展而建立起来的广泛的集合论基础。在实现方面的一个明显优点是,膨胀和腐蚀都属于基本操作,它们是各类形态学算法的基础。正如在接下来的章节中所显示的那样,形态学可以作为在广泛的应用领域中建立图像分割方法的基础。如在第 11 章中讨论的,形态学技术在图像描绘方法方面同样扮演着主要角色。

参考资料

Serra[1982]写的书是关于形态学图像处理的基础性参考资料。也可以查阅 Serra[1988],Giardina 和 Dougherty[1988]以及 Haralick 和 Shapiro[1992]的相关著述。其他与我们讨论内容相关的早期参考资料包括 Blum[1967],Lantuéjoul[1980],Maragos[1987]以及 Haralick 等人[1987]的材料。对于二值和灰度级形态学的综述可以参阅 Basart 和 Gonzalez[1992]以及 Basart 等人[1992]的著述。这组参考资料提供了 9.1 节到 9.4 节中涵盖的材料的丰富背景。

关于实现 9.5 节和 9.6 节中给出的这一类形态学算法的重要问题在 Jones 和 Svalbe[1994],Park 和 Chin[1995],Sussner 和 Ritter[1997],Anelli 等人[1998]以及 Shaked 和 Bruckstein[1998]的文章中都有说明。目前在形态学图像处理的理论和应用方面的研究工作总结在 Goutsias 和 Bloomberg[2000]写的书中,及名为 Pattern Recognition[2000]的特刊中。也可以参阅近期由 Rosenfeld[2000]编写的参考资料。由 Marchand-Maillet 和 Sharaiha[2000]写的关于二值图像处理的书及 Ritter 和 Wilson[2001]写的有关图像代数学的书也应给予关注。

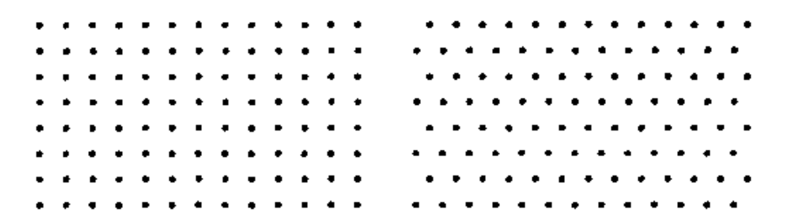
习题

9.1 本书中的数字图像嵌入了方形网格,像素可以是4,8或m连通的。然而,其他网格排列也是可能的。特殊情况下,有时也会使用具有6连通性的六边形网格排列(见下面的图形)。

(a)如何将图像从方形网格转换到六边形网格?

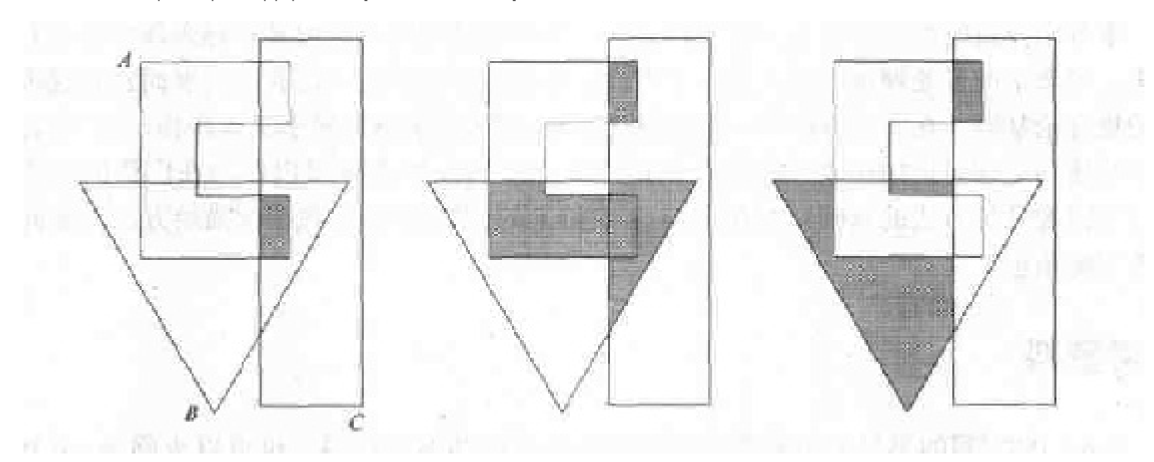
(b)对比讨论在方形网格中和六边形网格中对象旋转时的形状不变性。

(c)六边形网格是否会有与8连通情况下相同的多种结构(见2.5.2节)?



9.2 ★(a)参考图9.1(a),画出集合$(A\cap B)\cup(A\cup B)^c$的草图。

(b)用表达式写出下列图形中的阴影部分。每幅图像中的阴影区域构成一个集合,对3幅图像分别给出相应的表达式。



9.3 ★(a)给出将8连通二值边界转换成m连通边界的形态学算法(见2.5.2节)。可以假设边界是完全连通且为1个像素宽的。

(b)此算法的执行是否需要用每个结构元素进行多于一次的迭代过程?解释原因。

(c)算法的效能与使用结构元素的次序是否有关?如果是,请证明;否则举出例子说明此过程对于应用结构元素的次序的独立性。

9.4 只要B的原点包含于B中,则使用结构元素B对集合A进行腐蚀的结果是A的子集。给出腐蚀$A\ominus B$的结果位于A之外或部分位于A之外的例子。

9.5 下列4种说法是正确的。提出能确定这些说法正确的论据。(a)说法通常是正确的;(b)到(d)的说法只有对数字集合才是正确的。为了说明(b)到(d)说法的正确性,画一个离散的方形网格(如习题9.1中所显示的),并对每种情况使用网格上的点构成的集

合举出例了。提示:在能确定说法正确性的前提下尽量减少每种情况下点的数目。

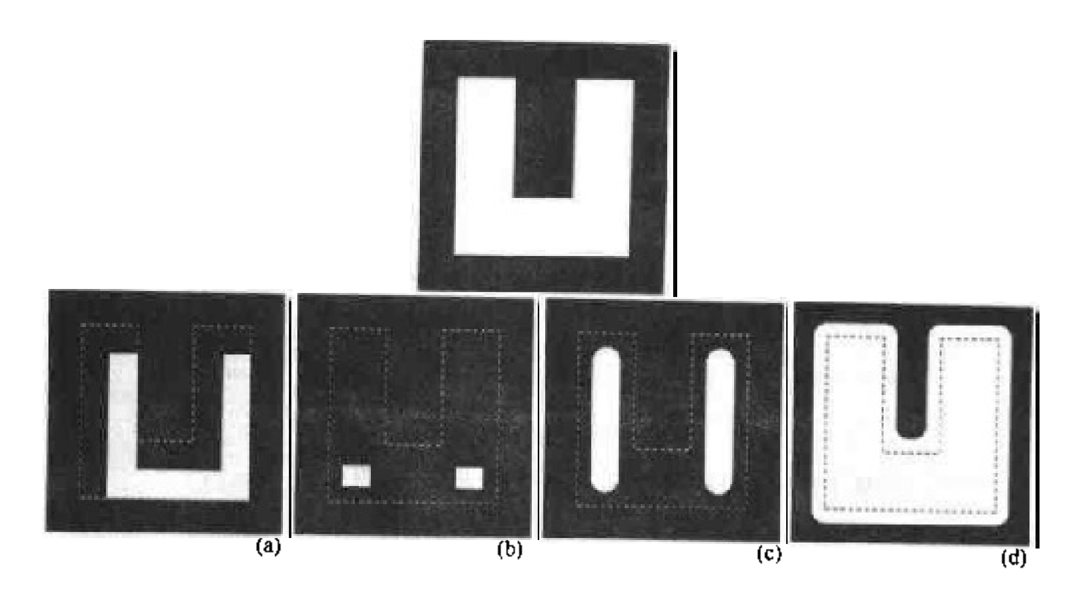
★(a)用凸结构元素对凸集的腐蚀结果还是凸集。

★(b)用凸结构元素对凸集的膨胀不一定总是凸形。

(c)数字凸集内的点不一定是连通的。

(d)存在这样的点的集合:集合中的每对点都有线相连,且线都位于集合内部,但这个集合不是凸形的。

★9.6 参考显示的图像,给出结构元素和形态学运算。运算的每一个结果显示在图(a)到(d)中。说明每个结构元素的原点。虚线部分说明了原始集合的边界,仅作为参考。注意在(d)中所有的拐角都是圆滑的。



9.7 令 A 表示图形中(见下页图)阴影表示的集合。参考显示的结构元素(黑点表示原点),画出下列形态学运算的结果:

(a)$(A \ominus B^4) \oplus B^2$

(b)$(A \ominus B^1) \oplus B^3$

(c)$(A \oplus B^1) \oplus B^3$

(d)$(A \oplus B^3) \ominus B^2$

★9.8 (a)反复进行图像膨胀操作的限制作用是什么?假设不使用通常的(一个点)结构元素。

(b)从(a)的答案出发,能够得到的最小图像是什么?

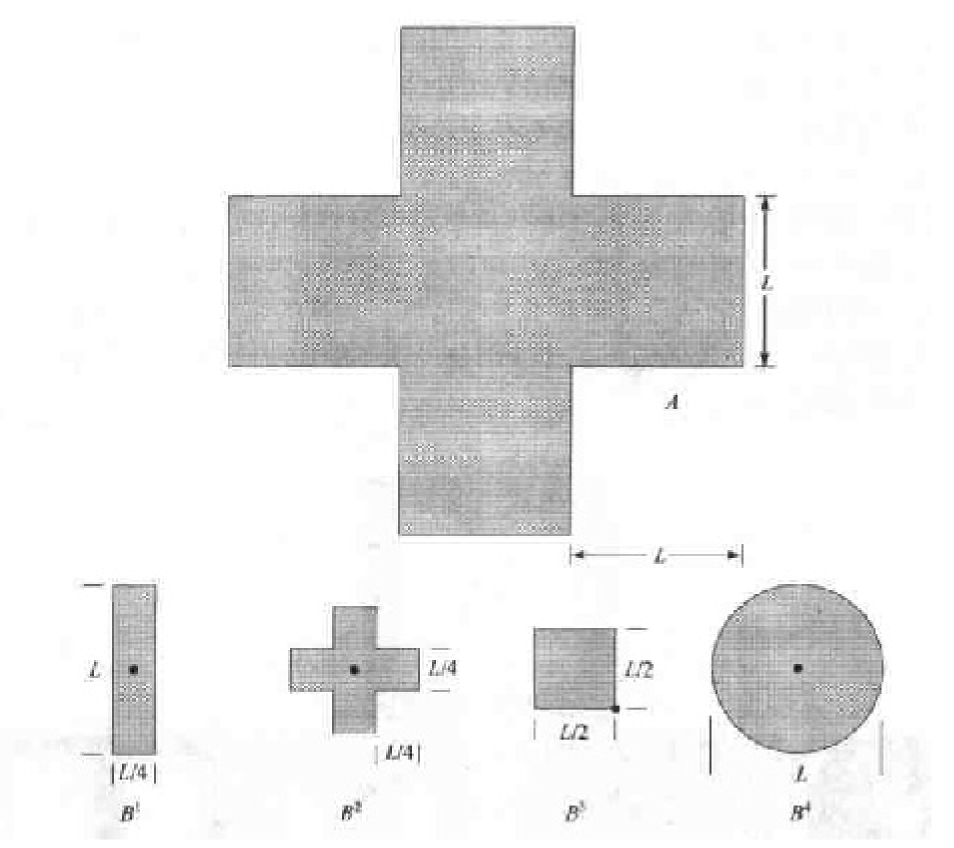
9.9 (a)反复进行图像腐蚀操作的限制作用是什么?假设不使用通常(一个点)的结构元素。

(b)从(a)的答案出发,能够得到的最小图像是什么?

★9.10 关于膨胀的另一个可替换的定义是:

$$A \oplus B = \{w \in Z^2 \mid w = a + b, a \in A, b \in B\}$$

说明该定义与式(9.2.1)的定义等价。



(习题9.7图)

9.11 (a)说明习题9.10中给出的定义和另一个膨胀定义等价:

$$A \oplus B = \bigcup_{b \in B} (A)_b$$

(这个表达式也叫做两个集合的闵可夫斯基和式。)

(b)说明(a)中的表达式同式(9.2.1)的定义是等价的。

★**9.12** 关于腐蚀的一个可替换的定义是:

$$A \ominus B = \{w \in Z^2 \mid w + b \in A, b \in B\}$$

说明这个定义与式(9.2.3)的定义是等价的。

9.13 (a)说明习题9.12中给出的腐蚀定义与另一个腐蚀定义是等价的:

$$A \ominus B = \bigcap_{b \in B} (A)_{-b}$$

(如果 $-b$ 用 b 代替,这个表达式称为两个集合的闵可夫斯基减法式。)

(b)说明(a)中的表达式与式(9.2.3)中给出的定义是等价的。

★**9.14** 证明二元表达式 $(A \cdot B)^c = (A^c \circ \hat{B})$ 的正确性。

9.15 证明下列表达式的正确性:

★(a)$A \circ B$ 是 A 的子集(子图)

(b)如果 C 是 D 的子集,则 $C \circ B$ 是 $D \circ B$ 的子集

(c)$(A \circ B) \circ B = A \circ B$

9.16 证明下列表达式的正确性[假设 B 的原点包含于 B 内,且习题9.16(a)为真]:

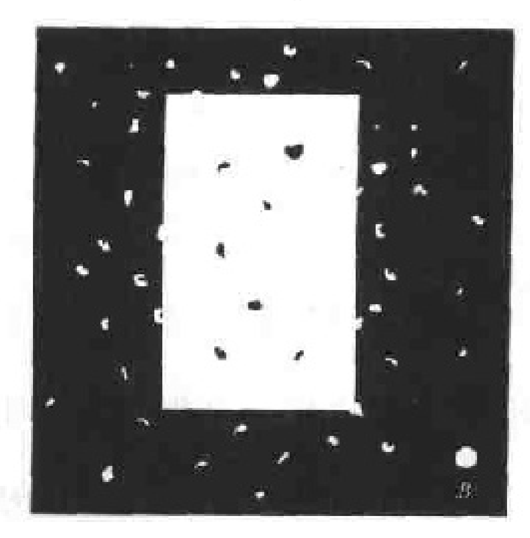
(a)A 是 $A \cdot B$ 的子集(子图)

(b)如果 C 是 D 的子集,则 $C \cdot B$ 是 $D \cdot B$ 的子集

(c) $(A \cdot B) \cdot B = A \cdot B$

9.17 参考所示的图像(下图)和结构元素,画出经下列操作后 C, D, E, F 是什么样子:$C = A \ominus B$; $D = C \oplus B$; $E = D \oplus B$; $F = E \ominus B$。初始集合 A 由除了结构元素 B 以外的所有白色图像分量构成。

注意:这个操作序列仅是用 B 对 A 进行开操作,而后对得到的结果用 B 进行闭操作。可以假设 B 的大小足够包围每个噪声成分。



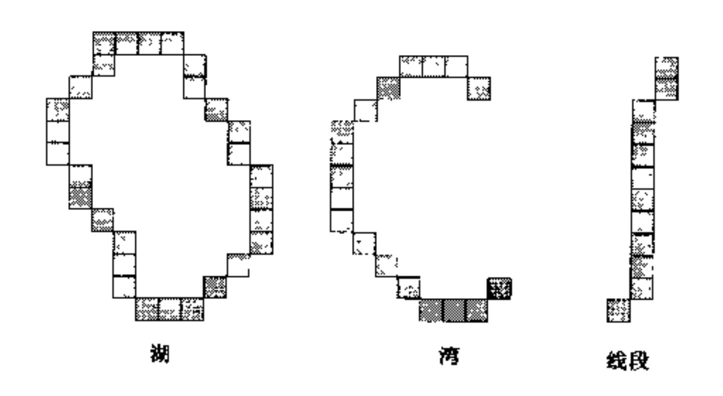
★9.18 在图9.7的例子中,腐蚀步骤中存留下来的正方形可以通过用于腐蚀的结构元素对这些正方形进行膨胀而得到恢复。膨胀之前做的腐蚀是对图像进行的开操作。一般来说,开操作不会将对象重构成原来的对象。解释为什么将图 9.7 中存留下来的正方形进行完全重构是可能的。

9.19 (a)画出对图像应用击中或击不中变换得到的结果和显示的结构元素。清楚地指出为结构元素选择的原点和边界。

(b)将得到的结果与用相关匹配的结果进行比较(图 4.46),说明两个结果之间的差异和相似点。



★9.20 区分图像中的细化对象时有用的三种特征(湖,湾和线段)示于下图中。建立一种在这些图形中可以进行区分的形态学逻辑算法。算法的输入是这三种图形中的一种。输出的是输入图形的识别。可以假设每种特征图形都是 1 个像素宽且是完全连通的。然而它们出现时的方向可以是任意的。



9.21 讨论在下列不同情况下，应该得到什么样的结果：

(a)对于9.5.2节中讨论的区域填充算法的起点是目标边界上的一点。

(b)区域填充算法的起点位于目标边界之外。

(c)画出使用9.5.4节中给出的算法计算的习题9.7的图形中凸壳的样子。假设$L=3$个像素。

★**9.22** (a)讨论使用图9.15(c)中给出的结构元素代替图9.13(b)中给出的结构元素提取边界后的效果。

(b)将元素值为1的3×3的结构元素代替图9.15(c)中的结构元素后，对式(9.5.2)的区域填充算法的影响是什么？

9.23 对图9.16中的例子提出一种完全自动化的方法。判定哪些黑色点是背景点和哪些点包含在球体之中(也就是，黑色区域完全包含在白色区域内)。假设二进制值1表示白色。

★**9.24** 9.5.3节中给出的提取连通分量的算法要求在每个连通分量中要有1个已知点以便提取整个连通分量的点。假如给出了包含任意数量(未知)的连通分量的二值图像，提出一种完全自动化的程序用于提取所有的连通分量。假设属于连通分量的点标记为1，背景点标记为0。

9.25 假设式(9.6.1)中的图像$f(x,y)$和结构元素$b(x,y)$都是矩形，域D_f和D_b分别表示$([F_{x1},F_{x2}],[F_{y1},F_{y2}])$和$([B_{x1},B_{x2}],[B_{y1},B_{y2}])$。例如，闭区间$[F_{x1},F_{x2}]$和$[F_{y1},F_{y2}]$是$xy$平面上沿$x$轴和$y$轴方向$x$和$y$的取值范围。函数$f(x,y)$在这个范围内有定义。

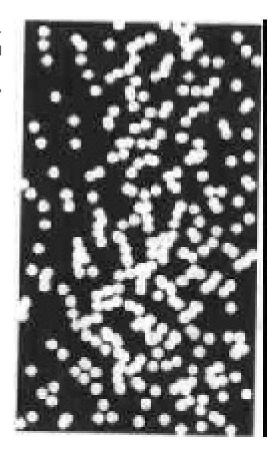
★(a)假设$(x,y)\in D_b$，推导对区间进行变换的表达式使替换变量s和t能满足式(9.6.1)。这个在s轴和t轴方向上的区间定义了st平面上$(f\oplus b)(s,t)$的矩形域。

(b)对式(9.6.2)定义的腐蚀重复(a)。

9.26 一幅灰度图像$f(x,y)$搀杂有彼此不重叠的噪声尖峰信号。这些噪声信号可以模拟为小的、半径为$R_{min}\leqslant r\leqslant R_{max}$、幅度为$A_{min}\leqslant a\leqslant A_{max}$的圆柱。

★(a)建立一种用于清理图像的形态学过滤方法。

(b)重复(a)，但假设最多有4个重叠的噪声尖峰信号。



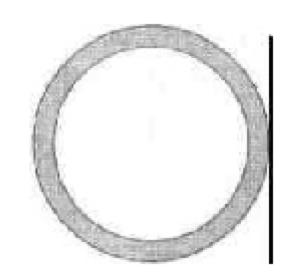
9.27 在显微镜方法应用方面的一个预处理是从两个或更多的交叠成堆的颗粒(见右图)中分离出单个独立的颗粒。假设所有颗粒具有相同的大小,提出一种形态学算法产生3幅图像,分别包含:

★(a)仅与图像边界融合在一起的颗粒。

(b)彼此交叠的颗粒。

(c)未交叠在一起的颗粒。

9.28 一种高技术生产工厂得到了政府的合同,制造一种如下图所示的高精度垫圈。合同的条文要求所有垫圈的形状要用一种图像系统检测。文中,形状检测是指检查垫圈的内外边缘是否偏离圆形。可进行下列假设:(1)对可接受垫圈的一幅"极好"的(对于这个问题来说是理想的)图像是可用的;(2)成像和定位系统在系统中的使用使系统具有的高精度足以忽略由于图像数字化过程和定位带来的误差。你作为顾问协助设计系统的视觉检测部分。提出一种以形态学或逻辑运算为基础的解决方案。将你的方案用方框图表示出来。



第10章 图像分割

前一章的资料使我们所研究的图像处理方法开始发生了转变。从输入输出均为图像的处理方法转变为,输入为图像而输出为从这些图像中提取出来的属性的处理方法(这方面在1.1节中定义过)。图像分割是这一方向的另一主要步骤。

分割将图像细分为构成它的子区域或对象。分割的程度取决于要解决的问题。就是说,在应用中,当感兴趣的对象已经被分离出来时,就停止分割。例如,在电子元件的自动检测方面,我们关注的是分析产品的图像,客观地检测是否存在特定的异常状态,比如,缺失的元件或断裂的连接线路。超过识别这些元件所需的分割是没有意义的。

异常图像的分割是图像处理中最困难的任务之一。精确的分割决定着计算分析过程的成败。因此,应该对提高稳定分割的可能性给予特别的关注。在某些情况下,比如工业检测应用,至少对环境进行适度控制的检测有时是可能的。有经验的图像处理系统设计师总是将相当大的注意力放在这类可能性上。在其他应用方面,比如自动目标采集,系统设计者无法对环境进行控制。所以,通常的方法是将注意力集中于传感器类型的选择上,这样可以增强获取所关注对象的能力,从而减少图像无关细节的影响。一个很好的例子就是,军方利用红外线图像发现有很强热信号的目标,比如移动中的装备和部队。

图像分割算法一般是基于亮度值的两个基本特性之一:不连续性和相似性。第1类性质的应用途径是基于亮度的不连续变化分割图像,比如图像的边缘。第2类的主要应用途径是依据事先制定的准则将图像分割为相似的区域。门限处理、区域生长、区域分离和聚合都是这类方法的实例。

本章中,我们将对刚刚提到的两类特性各讨论一些方法。我们先从适合于检测灰度级的不连续性的方法展开,如点、线和边缘。特别是边缘检测近年来已经成为分割算法的主题。除了边缘检测本身,我们还会讨论一些连接边缘线段和把边缘"组装"为边界的方法。关于边缘检测的讨论将在介绍了各种门限处理技术之后进行。门限处理也是一种人们普遍关注的用于分割处理的基础性方法,特别是在速度因素占重要地位的应用中。关于门限处理的讨论将在几种面向区域的分割方法展开的讨论之后进行。之后,我们将讨论一种称为分水岭分割法的形态学图像分割方法。这种方法特别具有吸引力,因为它将基于本章第一部分提到的技术的几种分割属性结合了起来。我们将以关于图像分割的动态提示应用方面的讨论来结束本章。

10.1 间断检测

在本节中,我们介绍几种用于检测数字图像中三种基本类型的灰度级间断的技术:点、线和边缘。寻找间断最一般的方法是,以3.5节中描述的方式对整幅图像使用一个模板进行检测。对一个图10.1所示的3×3模板,这一过程包括计算模板所包围区域内灰度级与模板系数的乘积之和。就是说,关于式(3.5.3),在图像中任意点的模板响应由下列公式给出:

$$R = w_1 z_1 + w_2 z_2 + \cdots + w_9 z_9 = \sum_{i=1}^{9} w_i z_i \qquad (10.1.1)$$

这里 z_i 是与模板系数 w_i 相联系的像素的灰度级。照例，模板的响应定义是相对于它的中心位置。有关执行模板操作的细节在3.5节中讨论。

w_1	w_2	w_3
w_4	w_5	w_6
w_7	w_8	w_9

图10.1　一个一般的3×3模板

10.1.1　点检测

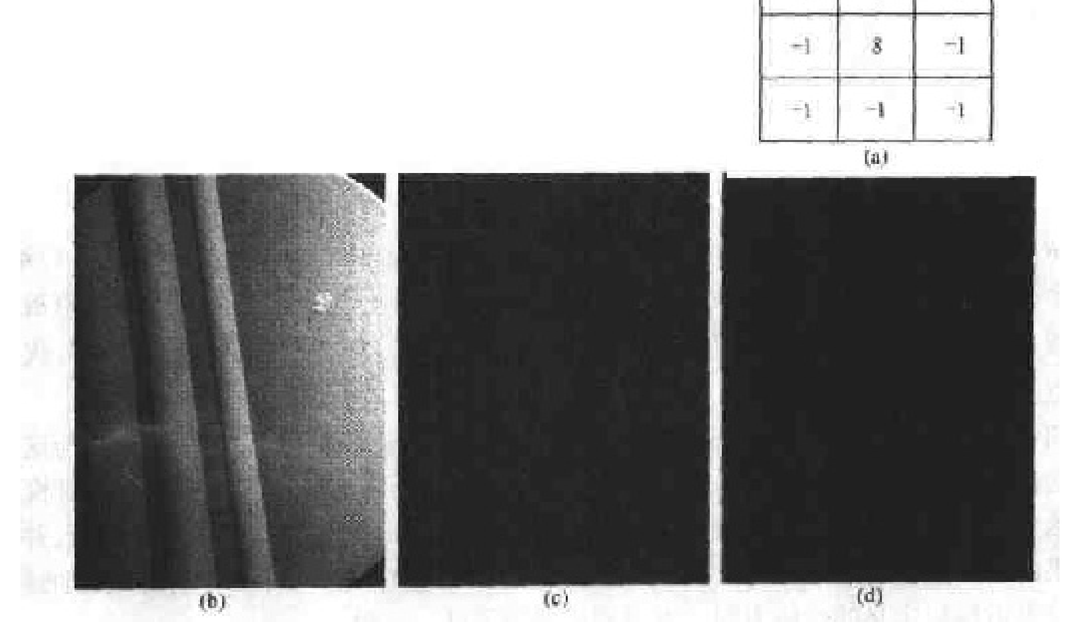
在一幅图像中，孤立点的检测在理论上是简单的。使用如图10.2(a)所示的模板，如果

$$|R| \geq T \qquad (10.1.2)$$

我们说在模板中心的位置上已经检测到一个点。这里 T 是一个非负门限，R 由式(10.1.1)给出。基本上，这个公式是测量中心点和它的相邻点之间加权的差值。基本思想就是：如果一个孤立的点(此点的灰度级与其背景的差异相当大并且它所在的位置是一个均匀的或近似均匀的区域)与它周围的点很不相同，则很容易被这类模板检测到。注意，图10.2(a)中的模板同图3.39(d)中给出的模板在拉普拉斯操作方面是相同的。严格地讲，这里强调的是点的检测。即，我们着重考虑的差别是那些足以识别为孤立点的差异(由 T 决定)。注意，模板系数之和为零表示在灰度级为常数的区域，模板响应为零。

-1	-1	-1
-1	8	-1
-1	-1	-1

(a)



(b)　(c)　(d)

图10.2　(a)点检测模板，(b)带有通孔的涡轮叶片的X射线图，(c)点检测的结果，(d)使用式(10.1.2)得到的结果(原图由X-TEK系统公司提供)

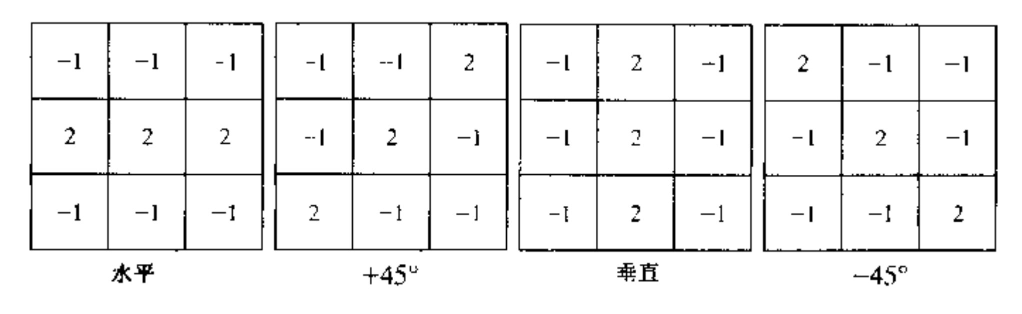
例10.1　图像中孤立点的检测

我们以图10.2(b)为辅助说明如何从一幅图中将孤立点分割出来，这幅X射线图显示了一个带有通孔的喷气发动机涡轮叶片，通孔位于图像的右上象限。在孔中只嵌有一个黑

色像素。图 10.2(c)是将点检测模板应用于 X 射线图像后得到的结果,图 10.2(d)显示了当 T 取图 10.2(c)中像素最高绝对值的 90%时,应用式(10.1.2)所得的结果(门限选择将在 10.3 节中详细讨论)。图中的这个单一的像素清晰可见(这个像素被人为放大以便印刷后可以看到)。由于这类检测是基于单像素间断,并且检测器模板的区域有一个均匀的背景,所以这个检测过程是相当有专用性的。当这一条件不能满足时,本章中讨论的其他方法会更适合检测灰度级间断。

10.1.2 线检测

复杂程度更高一级的检测是线检测。考虑图 10.3 中显示的模板。如果第 1 个模板在图像中移动,这个模板将对水平方向的线条(一个像素宽度)有更强的响应。在一个不变的背景上,当线条经过模板的中间一行时会产生响应的最大值。画一个元素为 1 的简单阵列,并且使具有不同灰度级(如 5)的一行水平穿过阵列,可以很容易验证这一点。同样的实验可以显示出图 10.3 中的第 2 个模板对于 45°方向线有最佳响应;第 3 个模板对于垂直线有最佳响应;第 4 个模板对于 -45°方向线有最佳响应;这些方向也可以通过注释每个模板的优选方向来设置,即,在这些方向上用比别的方向更大的系数(为 2)设置权值。注意每个模板系数相加的总和为零,表示在灰度级恒定的区域来自模板的响应为零。



水平

-1	-1	-1
2	2	2
-1	-1	-1

+45°

-1	-1	2
-1	2	-1
2	-1	-1

垂直

-1	2	-1
-1	2	-1
-1	2	-1

-45°

2	-1	-1
-1	2	-1
-1	-1	2

图 10.3 线模板

令 R_1, R_2, R_3 和 R_4 从左到右代表图 10.3 中模板的响应,这里 R 的值由式(10.1.1)给出。假设 4 个模板分别应用于一幅图像,在图像中心的点,如果 $|R_i| > |R_j|, j \neq i$,则此点被认为与在模板 i 方向上的线更相关。例如,如果在图中的一点有 $|R_1| > |R_j|, j = 2,3,4$,我们说此特定点与水平线有更大的联系。

换句话说,我们可能对检测特定方向上的线感兴趣。在这种情况下,我们应使用与这一方向有关的模板,并设置该模板的输出门限,如式(10.1.2)所示。换句话说,如果我们对检测图像中由给定模板定义的方向上的所有线感兴趣,只需要简单地通过整幅图像运行模板,并对得到的结果的绝对值设置门限即可。留下的点是有最强响应的点。对于一个像素宽度的线,这些响应最靠近模板定义的对应方向。下列例子说明了这一过程。

例 10.2 特定方向上的线检测

图 10.4(a)显示了一幅电路接线模板的数字化(二值的)图像。假设我们要找到一个像素宽度的并且方向为 -45°的线条。基于这个假设,使用图 10.3 中最后一个模板。图 10.4(b)显示了得到的结果的绝对值。注意,图像中所有水平和垂直的部分都被除去了,并且在图 10.4(b)中所有原图中接近 -45°方向的部分产生了最强响应。

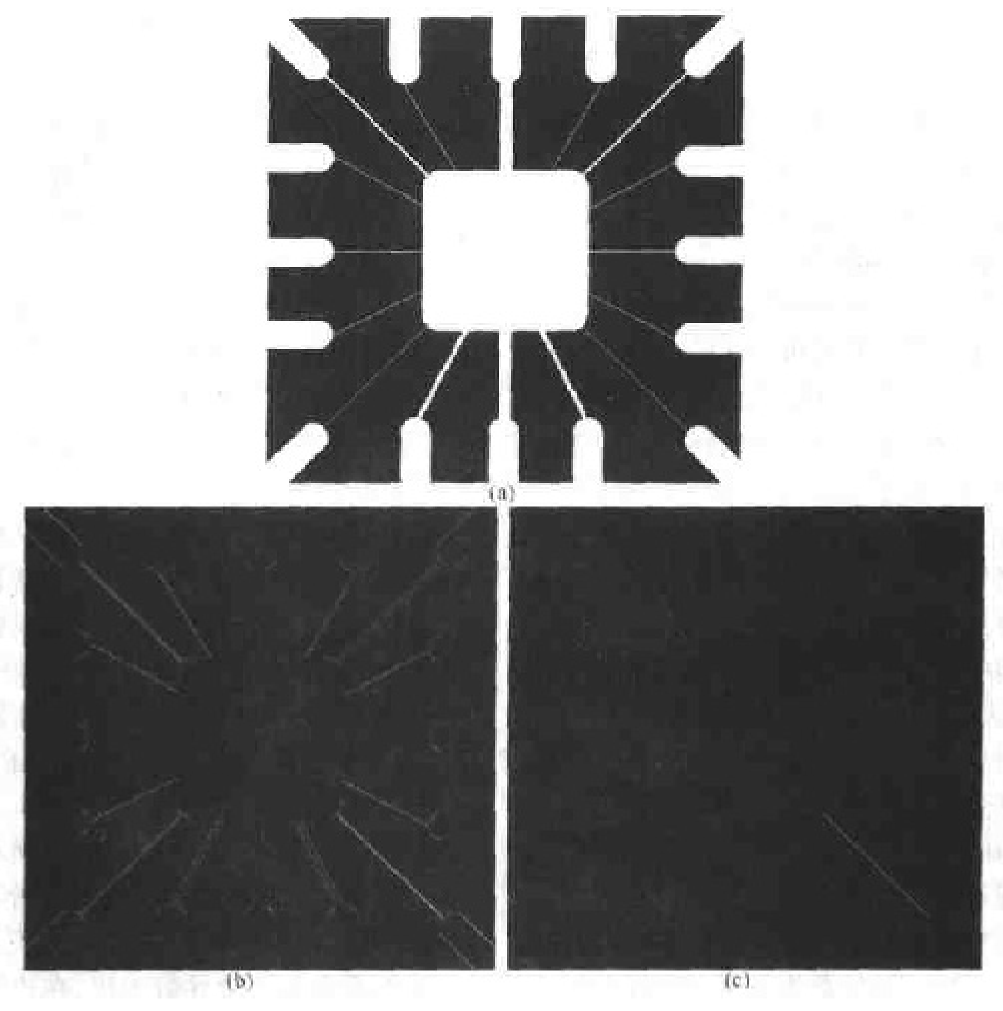


图 10.4 线检测的说明。(a)二值电路接线模板,(b)使用 - 45°线检测器处理后得到的绝对值,(c)对图像(b)设置门限得到的结果

为了决定哪一条线拟合模板最好,只需要简单地对图像设置门限。图10.4(c)显示了使门限等于图像中最大值后得到的结果。对于与这个例子类似的应用,让门限等于最大值是一个好的选择,因为输入图像是二值的,并且我们要寻找的是最强响应。图10.4(c)显示了在白色区所有通过门限检测的点。此时,这一过程只提取了一个像素宽且方向为 - 45°的线段(图像中在左上象限中也有此方向上的图像部分,但宽度不是一个像素)。图 10.4(c)中显示的孤立点是对于模板也有相同强度响应的点。在原图中,这些点和与它们紧挨着的相邻点,是用模板在这些孤立位置上生成最大响应的方法来定向的。这些孤立点也可以使用图 10.2(a)中的模板进行检测,然后删除,或者使用下一章中讨论的形态学腐蚀法删除。

10.1.3 边缘检测

尽管在任何关于分割的讨论中,点和线检测都是很重要的,但是边缘检测对于灰度级间断的检测是最为普遍的检测方法。本节中,我们讨论实现一阶和二阶数字导数检测一幅图像中边缘的方法。在 3.7 节介绍图像增强的内容中介绍过这些导数。本节的重点将放在边缘检测的特性上。某些前面介绍的概念在这里为了叙述的连续性将进行简要的重述。

基本说明

在3.7.1节中我们非正式地介绍过边缘。本节中我们更进一步地了解数字化边缘的概念。直观上,一条边缘是一组相连的像素集合。这些像素位于两个区域的边界上。然而,我们已经在2.5.2节中用一定的篇幅解释了一条边缘和一条边界的区别。从根本上讲,如我们将要看到的,一条边缘是一个"局部"概念,而由于其定义的方式,一个区域的边界是一个更具有整体性的概念。给边缘下一个更合理的定义需要具有以某种有意义的方式测量灰度级跃变的能力。

我们先从直观上对边缘建模开始。这样做可以将我们引领至一个能测量灰度级有意义的跃变的形式体系中。从感觉上说,一条理想的边缘具有如图10.5(a)所示模型的特性。依据这个模型生成的完美边缘是一组相连的像素的集合(此处为在垂直方向上),每个像素都处在灰度级跃变的一个垂直的台阶上(如图形中所示的水平剖面图)。

实际上,光学系统、取样和其他图像采集的不完善性使得到的边缘是模糊的,模糊的程度取决于诸如图像采集系统的性能、取样率和获得图像的照明条件等因素。结果,边缘被更精确地模拟成具有"类斜面"的剖面,如图10.5(b)所示。斜坡部分与边缘的模糊程度成比例。在这个模型中,不再有细线(一个像素宽的线条)。相反,现在边缘的点是包含于斜坡中的任意点,并且边缘成为一组彼此相连接的点集。边缘的"宽度"取决于从初始灰度级跃变到最终灰度级的斜坡的长度。这个长度又取决于斜度,斜度又取决于模糊程度。这使我们明白:模糊的边缘使其变粗而清晰的边缘使其变得较细。

图10.6(a)显示的图像是从图10.5(b)的放大特写中提取出来的。图10.6(b)显示了两个区域之间边缘的一条水平的灰度级剖面线。这个图形也显示出灰度级剖面线的一阶和二阶导数。当我们沿着剖面线从左到右经过时,在进入和离开斜面的变化点,一阶导数为正。在灰度级不变的区域一阶导数为零。在边缘与黑色一边相关的跃变点二阶导数为正,在边缘与亮色一边相关的跃变点二阶导数为负,沿着斜坡和灰度为常数的区域为零。在图10.6(b)中导数的符号在从亮到暗的跃变边缘处取反。

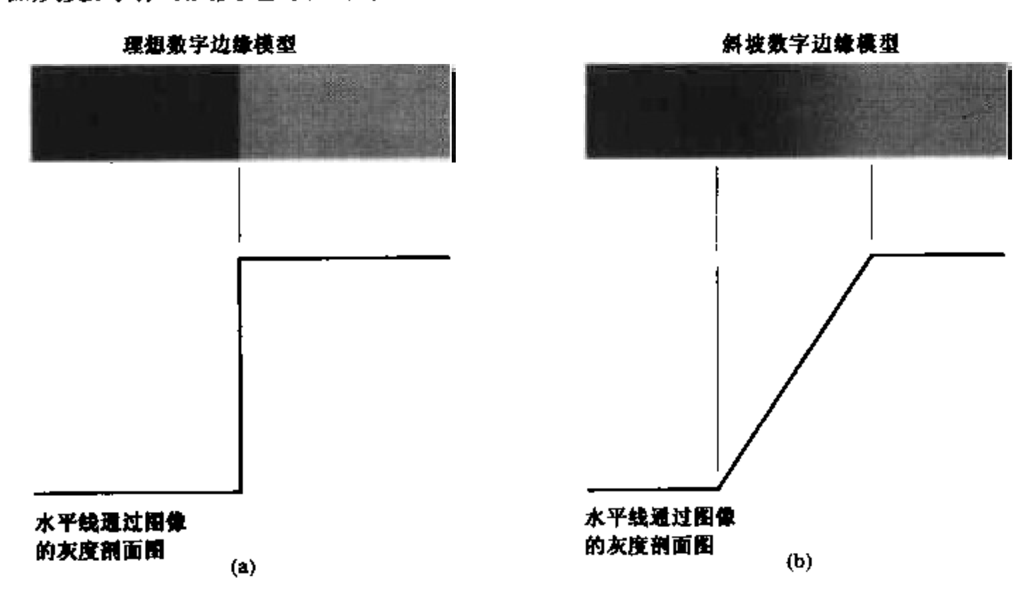


图10.5 (a)理想的数字边缘模型,(b)斜坡数字边缘模型。斜坡部分与边缘的模糊程度成比例

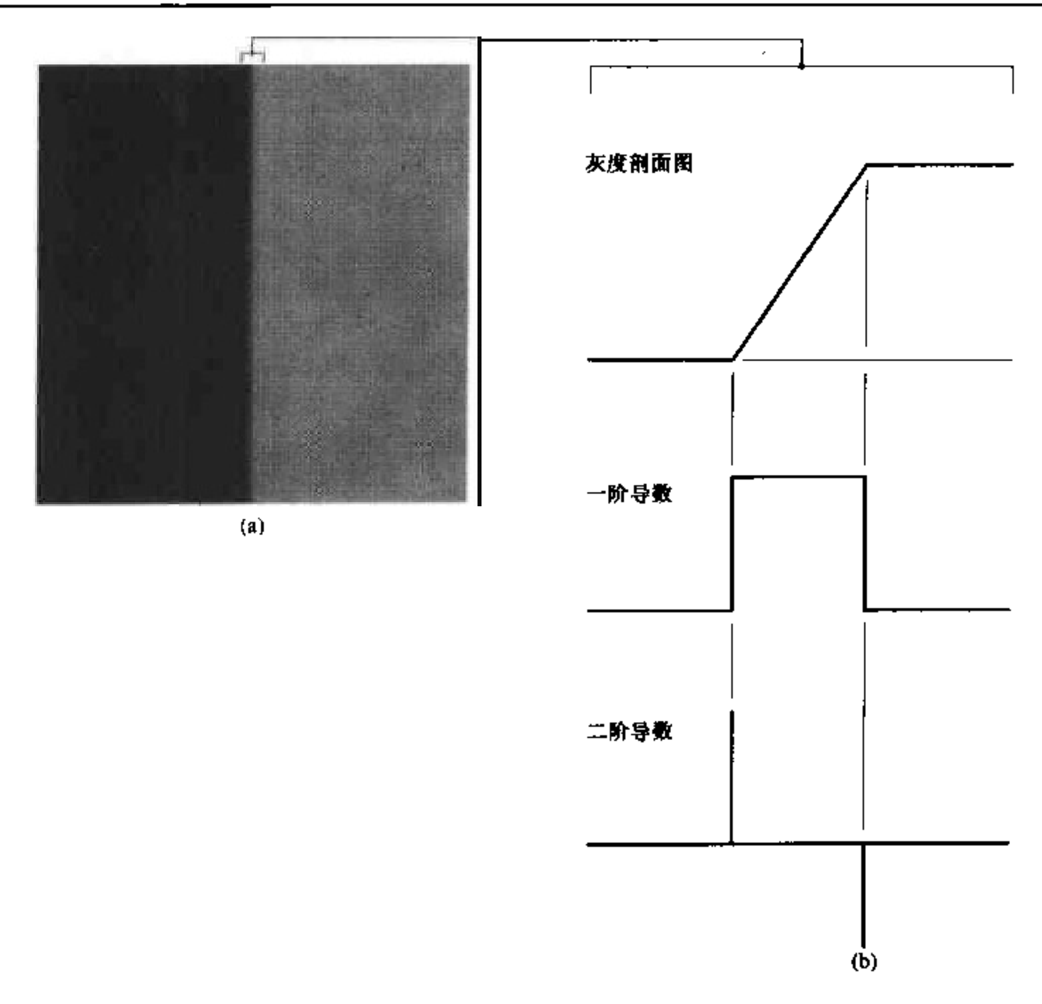


图 10.6 (a)由一条垂直边缘分开的两个不同区域,(b)边界附近的细节,显示了一个灰度级剖面图和一阶与二阶导数的剖面图

由这些现象我们可以得到的结论是:一阶导数可以用于检测图像中的一个点是否是边缘的点(也就是判断一个点是否在斜坡上)。同样,二阶导数的符号可以用于判断一个边缘像素是在边缘亮的一边还是暗的一边。我们注意到围绕一条边缘,二阶导数的两条附加性质:(1)对图像中的每条边缘二阶导数生成两个值(一个不希望得到的特点);(2)一条连接二阶导数正极值和负极值的虚构直线将在边缘中点附近穿过零点。将在本节后面说明,二阶导数的这个过零点的性质对于确定粗边线的中心非常有用。

最后,注意到某些边缘模型利用了在进入和离开斜坡地方的平滑过渡(习题10.5)。然而,我们在接下来的讨论中将得出同样的结论。而且,这一点从我们使用局部检测进行处理就可以很明显地看出(因此,2.5.2 节中对于边缘的局部性质进行了说明)。

尽管到此为止我们的注意力被限制在一维水平剖面线范围内,但同样的结论可以应用于图像中的任何方向上。我们仅仅定义了一条与任何需要考察的点所在的边缘方向相垂直的剖面线,并如前面讨论的那样,对结果进行了解释。

例 10.3 在有噪声的边缘附近的一阶和二阶导数性质

图 10.5 和图 10.6 所示的边缘是无噪声的。图 10.7 中第 1 列的图像分割显示了分割左右

黑白区域的4个斜坡边缘的特写图。这对于在头脑中形成整个从黑到白的跃变是一条单边缘的观点是很重要的。左上角的图像分割是无噪声的。图10.7第1列的其他3幅图分别被附加的零均值且标准差为0.1,1.0和10.0灰度级的高斯噪声污染了。每幅图像下面显示的图是穿过图像的水平扫描线的灰度级剖面线。

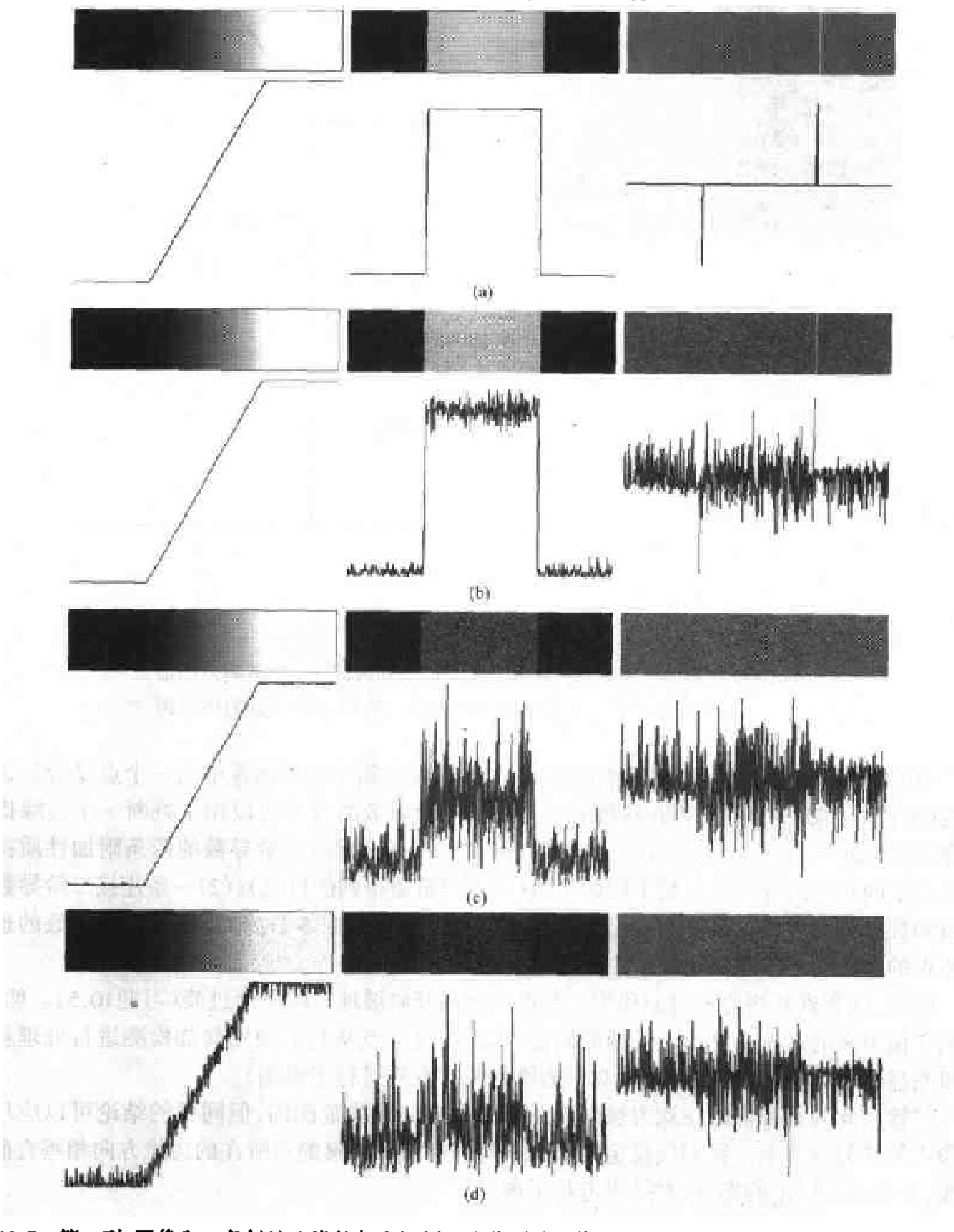


图10.7 第1列:图像和一条斜坡边缘的灰度级剖面线分别被均值为0且$\sigma=0.0,0.1,1.0$和10.0的随机高斯噪声污染了,第2列:一阶导数图像和灰度级剖面线,第3列:二阶导数图像和灰度级剖面线

图10.7的第2列图像是左边图像的一阶导数图像(将在下一节中讨论一阶和二阶图像导数)。例如,考虑顶部的中心图像。如同我们在讨论图10.6(b)时一样,在恒定的黑色和白色区域导数为零。这是在导数图像中的两个黑色区域。不变化的斜坡导数是常量,大小等于斜坡的斜率。这个在导数图像中不变的区域用灰色表示。当我们将中心列向下移时,导数变得与无噪声情况时越来越不相同。实际上,导数的变化使得更难将此列结尾处的剖面线与一个斜坡边缘联系起来。造成这种有趣结果的原因是,左列的图像中噪声几乎是不可见的了。最后一幅图像呈现轻微的颗粒状,但这对图像的破坏是极其细微的。这些例子很好地说明了导数对于噪声的敏感性。

正如我们预想的,二阶导数对于噪声甚至更为敏感。无噪声图像的二阶导数是显示在右上角的图像。浅黑色和白色的线是如图10.6所描述的正和负的分量。这些图像中的灰色描绘了由于比例缩放而生成的零点。我们注意到类似于无噪声情况的有噪声的二阶导数,是对应于灰度级为0.1的标准差的噪声。另两幅二阶导数图像和剖面线清楚地说明了,实际检测这些图像中为正和为负的分量很困难,而这些分量是在边缘检测中非常有用的二阶导数特性。

微小的噪声对在图像中检测边缘所用的两个主要导数有重要影响的事实,是应考虑的重要问题。特别是图像平滑处理时,在我们刚刚讨论的存在较小水平噪声的应用中,使用导数之前应该认真考虑。

基于这个例子和以前的3段论述,我们可以得出这样的结论:为了对有意义的边缘点进行分类,与这个点相联系的灰度级变换必须比在这一点的背景上的变换更为有效。由于我们用局部计算进行处理,决定一个值是否有效的选择方法就是使用门限。因此,如果一个点的二维一阶导数比指定的门限大,我们就定义图像中的此点是一个边缘点。一组这样的依据事先定好的连接准则(见2.5.2节)相连的边缘点就定义为一条边缘。术语"边缘线段"一般在边缘与图像的尺寸比起来很短时才使用。如10.2节中所阐明的,分割的关键问题是如何将边缘线段组合成更长的边缘。如果我们选择使用二阶导数,则另一个可用的定义是将图像中的边缘点定义为它的二阶导数的零交叉点。此时,边缘的定义同上面讲过的定义是一样的。应该注意到,这些定义并不能保证在一幅图像中成功地找到边缘。它们只是给了我们一个寻找边缘的形式体系。如第3章中所讲的,图像中的一阶导数用梯度计算。二阶导数使用拉普拉斯算子得到。

梯度算子

一幅数字图像的一阶导数是基于各种二维梯度的近似值。图像$f(x,y)$在位置(x,y)的梯度定义为下列向量:

$$\nabla \mathbf{f} = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f}{\partial x} \\ \dfrac{\partial f}{\partial y} \end{bmatrix} \tag{10.1.3}$$

从向量分析中我们知道,梯度向量指向在坐标(x,y)的f的最大变化率方向。

在边缘检测中,一个重要的量是这个向量的大小,用∇f表示,这里,

$$\nabla f = \mathrm{mag}(\nabla \mathbf{f}) = [G_x^2 + G_y^2]^{1/2} \tag{10.1.4}$$

这个量给出了在$\nabla \mathbf{f}$方向上每增加单位距离后$f(x,y)$值增大的最大变化率。一般来讲也将∇f

称为梯度(尽管并不完全正确)。我们也将坚持惯例,交替地使用这一术语,只在有可能混淆的地方将向量和它的大小区分开。

梯度向量的方向也是一个重要的量。令 $\alpha(x,y)$ 表示向量 $\nabla \mathbf{f}$ 在 (x,y) 处的方向角。然后,由向量分析得到:

$$\alpha(x,y) = \arctan\left(\frac{G_y}{G_x}\right) \tag{10.1.5}$$

这里,角度是以 x 轴为基准度量的。边缘在 (x,y) 处的方向与此点的梯度向量的方向垂直。

计算图像的梯度要基于在每个像素位置都得到了偏导数 $\partial f/\partial x$ 和 $\partial f/\partial y$。令图10.8(a)中显示的3×3大小的区域表示图像邻域中的灰度级。如在3.7.3节中讨论的,得到 z_5 点处的一阶偏导数的最简单方法之一是使用下列 Roberts 交叉梯度算子:

$$G_x = (z_9 - z_5) \tag{10.1.6}$$

和

$$G_y = (z_8 - z_6) \tag{10.1.7}$$

通过用图 10.8(b)中所示的模板及 3.5 节讨论的过程可以得到整幅图像的导数。

2×2 大小的模板由于没有清楚的中心点所以很难使用。使用 3×3 大小的模板的方法由下式给出:

$$G_x = (z_7 + z_8 + z_9) - (z_1 + z_2 + z_3) \tag{10.1.8}$$

和

$$G_y = (z_3 + z_6 + z_9) - (z_1 + z_4 + z_7) \tag{10.1.9}$$

在这组公式中,3×3 大小的图像区域的第 1 行和第 3 行间的差近似于 x 方向上的导数,第 3 列和第 1 列之差近似于 y 方向上的导数。图 10.8(d)和(e)中显示的模板称为Prewitt算子,可以用于计算这两个公式。

对这两个公式的一个小小的变化是在中心系数上使用一个权值 2:

$$G_x = (z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3) \tag{10.1.10}$$

和

$$G_y = (z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7) \tag{10.1.11}$$

权值 2 用于通过增加中心点的重要性而实现某种程度的平滑效果(习题 10.8)。图 10.8(f)和(g)称为Sobel算子,用以实现这两个公式。Prewitt和Sobel算子是在实践中计算数字梯度时最常用的。Prewitt模板实现起来比Sobel模板更为简单,但后者在噪声抑制特性方面略胜一筹,这在处理导数时是个重要的问题。注意,在图 10.8 中所有模板中的系数总和为 0,表示正如导数算子所预示的,此时在灰度级不变的区域,模板响应为 0。

刚才讨论过的模板是用于求梯度分量 G_x 和 G_y 的。梯度的计算需要这两个分量按式(10.1.4)显示的方式联合使用。然而,实际执行中并不总能令人满意。因为计算平方和平方根需要大量计算。经常使用的一种方法是用绝对值对梯度进行近似:

$$\nabla f \approx |G_x| + |G_y| \tag{10.1.12}$$

这个公式在计算方面更有吸引力,而且它仍保持着同灰度级的相对变化。如 3.7.3 节中提到的,得到这个优点的代价是通常导致滤波器不是各向同性的(对于旋转变换不变)。然而,当用Prewitt和Sobel一类的模板去计算 G_x 和 G_y 时这并不构成问题。只有对垂直和水平边缘这些模

板才会得出各向同性的结果,所以,即使我们使用式(10.1.4)计算梯度,只有在这些方向上的边缘结果才是各向同性的。此时,式(10.1.4)和式(10.1.12)给出相同的结果(习题10.6)。

可以调整图10.8中的3×3模板,以便它们在沿对角线方向上能有最大的响应。另外,两个用于检测对角线方向上突变的Prewitt和Sobel模板在图10.9中给出。

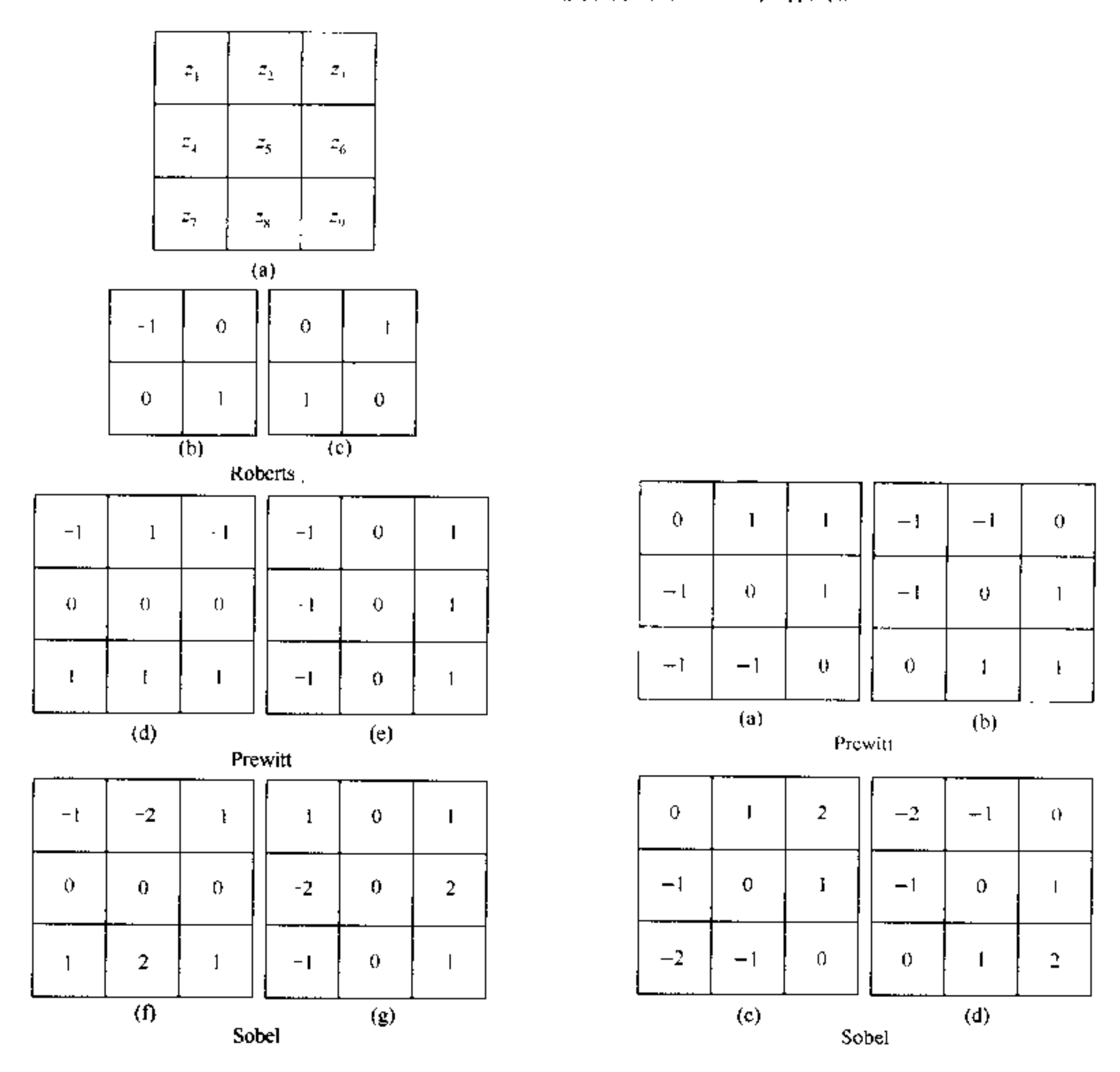


图10.8 一幅图像的3×3大小的区域(z是灰度值)和用于计算标记为z_5的点的梯度的不同模板

图10.9 用于检测对角边缘的Prewitt和Sobel模板

例10.4 梯度和它的分量说明

图10.10说明了梯度的两个分量$|G_x|$和$|G_y|$的响应与这两个分量之和生成的梯度图像。图10.10(b)和(c)中两个分量的方向性是很明显的。特别注意,图10.10(b)中屋瓦、砖块的水平接缝和窗户的水平分段的图像是非常清晰的。与之形成对照的是图10.10(c)中表现出的垂直部分,诸如墙附近的拐角、窗户的垂直部分、砖块的垂直接缝和图片右侧的街灯柱子。原图分辨率相当高(1 200×1 600像素),且在图片摄取的距离上,墙砖对于图像细节的影响仍十分显著。这种程度的细节通常是不符合要求的。减少这种影响的一种方法是对图像进行平滑处理。图10.11显示了图10.10中相同的图像序列,但对原图像首次使用了

一个 5×5 的均值滤波器进行了平滑处理。现在每个模板的响应几乎未显示出由砖块造成的影响,得到的结果几乎都是主要的边缘。注意取均值造成所有的边缘响应都被削弱了。

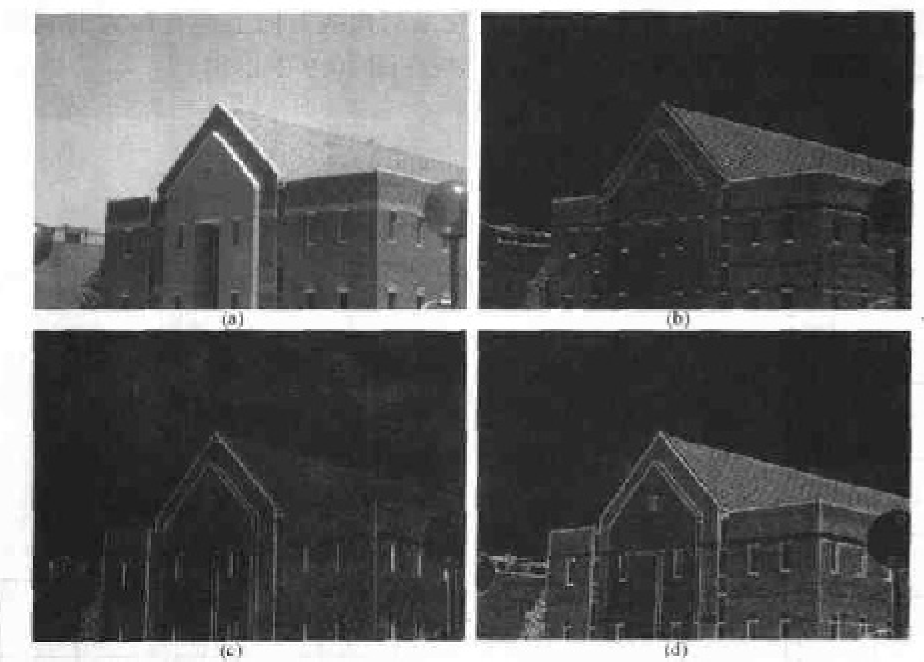


图 10.10 (a)原图,(b)$|G_x|$,x 方向上的梯度分量,(c)$|G_y|$,y 方向上的梯度分量,(d)梯度图像 $|G_x|+|G_y|$

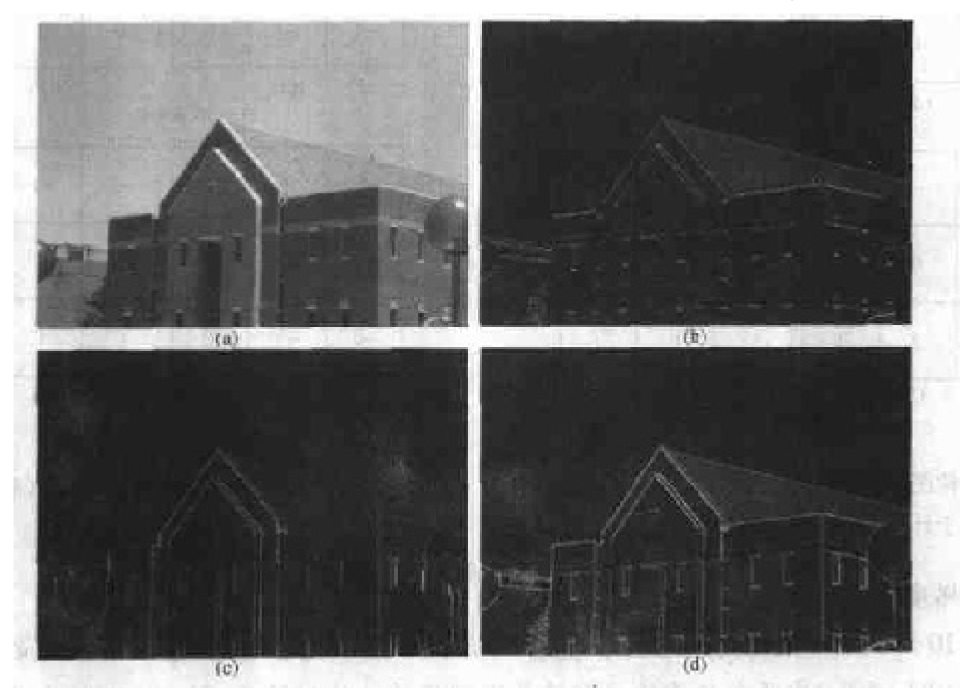


图 10.11 与图 10.10 相同的图像序列,但原图经过一个 5×5 的均值滤波器进行了平滑处理

在图 10.10 和图 10.11 中明显显示出,水平和垂直 Sobel 模板对正负 45°方向边缘的反映几乎一样好。如果这一点对于强调对角线方向十分重要,那么就会使用图 10.9 中的各个模板对中的一个。图 10.12 中显示的是对角Sobel模板的绝对响应。在这幅图中,这些模板更强的对角响应是很明显的。两个对角模板对水平和垂直的边缘具有相似的响应,但正

如所期望的那样，它们在这些方向上的响应比图 10.10(b)和图 10.10(c)中显示的Sobel模板在水平和垂直方向上的响应要弱。

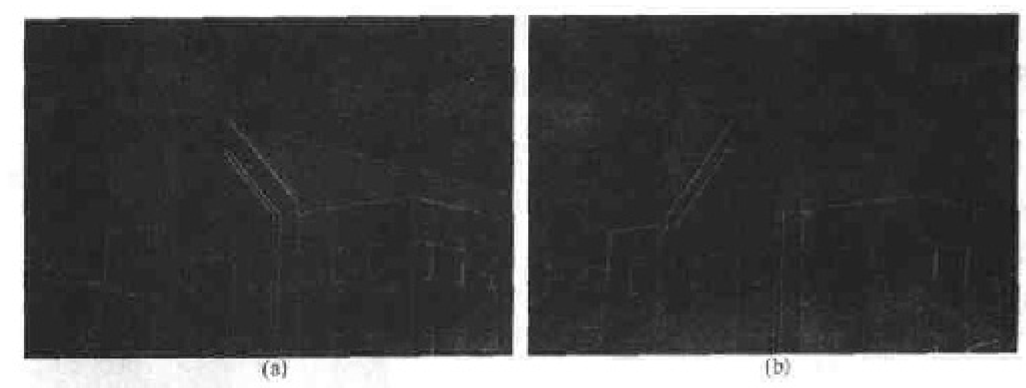


图 10.12　对角边缘检测。(a)用图10.9(c)模板的结果，(b)用图 10.9(d)模板的结果，两种情况的输入都是图 10.11(a)

拉普拉斯算子

二维函数 $f(x,y)$的拉普拉斯算子是如下定义的二阶导数：

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{10.1.13}$$

对拉普拉斯算子的数字近似方法在 3.7.2 节中定义过。对一个 3×3 大小的区域，在实践中经常遇到的两种形式之一是：

$$\nabla^2 f = 4z_5 - (z_2 + z_4 + z_6 + z_8) \tag{10.1.14}$$

这里 z 值在图 10.8 (a)中定义。包括对角邻域的数字近似方法由下式给出：

$$\nabla^2 f = 8z_5 - (z_1 + z_2 + z_3 + z_4 + z_6 + z_7 + z_8 + z_9) \tag{10.1.15}$$

计算这两个公式的模板示于图 10.13。我们由这些模板注意到式(10.1.14)和式(10.1.15)的计算对于分别以 90°和 45°为增量的旋转变换是各向同性的。

0	-1	0
-1	4	-1
0	-1	0

-1	-1	-1
-1	8	-1
-1	-1	-1

图 10.13　用于分别实现式(10.1.14) 和式(10.1.15)的拉普拉斯算子模板

拉普拉斯算子一般不以其原始形式用于边缘检测是由于存在下列原因：作为一个二阶导数，拉普拉斯算子对噪声具有无法接受的敏感性(图 10.7)；拉普拉斯算子的幅值产生双边缘(见图 10.6 和图 10.7)，这是复杂的分割不希望有的结果；最后，拉普拉斯算子不能检测边缘的方向。由于以上原因，拉普拉斯算子在分割中所起的作用包括：(1)如在本节前面提到的，利用它的零交叉的性质进行边缘定位，或(2)如我们在 10.3.6 节中说明的，确定一个像素是在一条边缘暗的一边还是亮的一边。

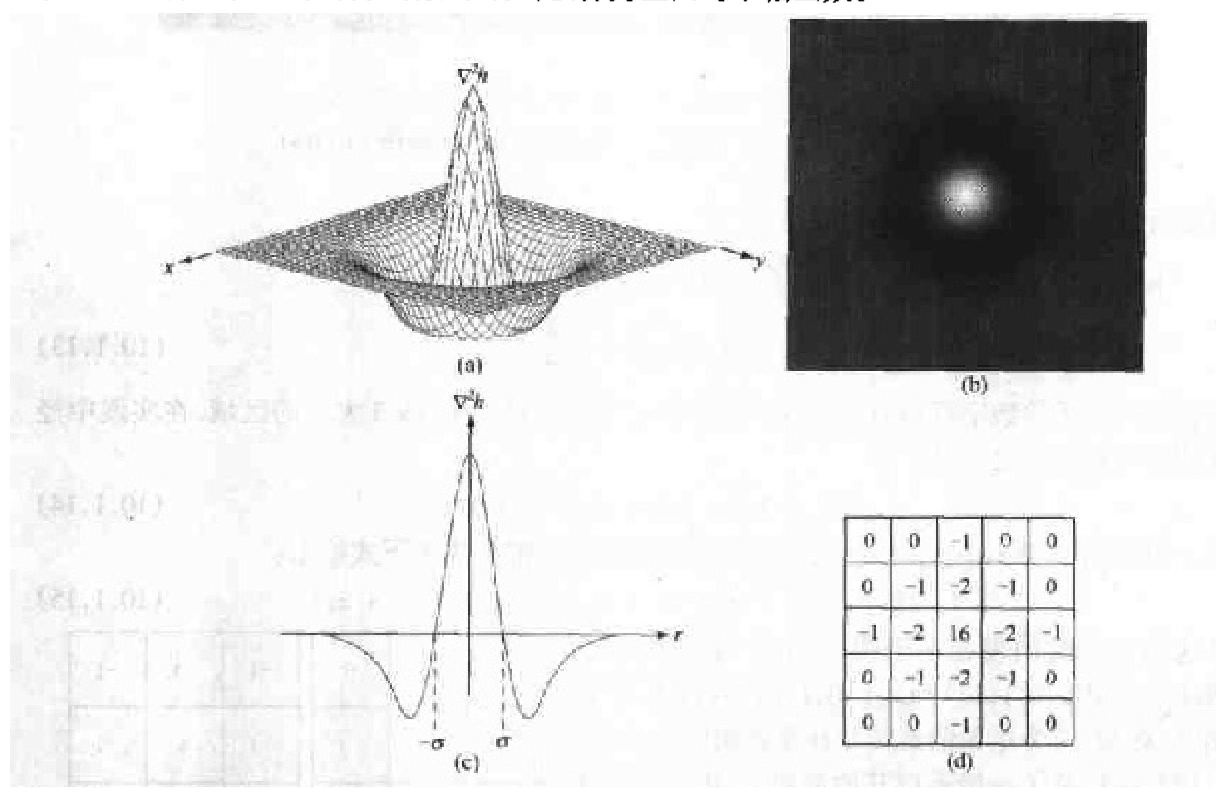
对第 1 类作用，拉普拉斯算子与平滑过程一起利用零交叉作为找到边缘的前兆。考虑函数：

$$h(r) = -e^{-\frac{r^2}{2\sigma^2}} \tag{10.1.16}$$

这里 $r^2 = x^2 + y^2$，σ 是标准差。用一幅图像与该函数卷积模糊该图像，图像模糊的程度是由 σ 值决定的。h 的拉普拉斯算子(h 关于 r 的二阶导数)是：

$$\nabla^2 h(r) = -\left[\frac{r^2-\sigma^2}{\sigma^4}\right]e^{-\frac{r^2}{2\sigma^2}} \tag{10.1.17}$$

这个公式一般叫做高斯型的拉普拉斯算子(Laplacian of a Gaussian,LoG),因为式(10.1.16)采用了高斯型函数的形式。图 10.14 显示了一幅LoG函数的三维曲线、图像和LoG函数的横截面。还显示了一个对$\nabla^2 h$近似的 5×5 模板。这种近似不是惟一的。其目的是得到$\nabla^2 h$ 本质的形状;即,一个正的中心项,周围被一个相邻的负值区域围绕(这个负值区域从原点开始作为距离的函数在值上是增加的),并被一个零值的外部区域所包围。系数的总和也必须为零,以便在灰度级不变的区域中模板的响应为零。这个小的模板仅对基本上无噪声的图像有用。由于图像的形状,高斯型的拉普拉斯算子有时被称为墨西哥草帽函数。



0	0	-1	0	0
0	-1	-2	-1	0
-1	-2	16	-2	-1
0	-1	-2	-1	0
0	0	-1	0	0

图 10.14 高斯型的拉普拉斯算子(LoG),(a)三维曲线,(b)图像(黑色是负值区域,灰色是零值平面,白色是正值区域),(c)零交叉的横截面显示,(d)图形(a)近似的5×5模板

因为二阶导数是线性运算,所以用$\nabla^2 h$ 卷积一幅图像与首先使用式(10.1.16)的高斯型平滑函数卷积该图像,然后计算所得结果的拉普拉斯是一样的。因此,我们看到在LoG公式中使用高斯型函数的目的就是对图像进行平滑处理,使用拉普拉斯算子的目的是提供一幅用零交叉确定边缘位置的图像。图像的平滑处理减少了噪声的影响并且它的主要作用还是抵消由拉普拉斯算子的二阶导数引起的逐渐增加的噪声影响。值得注意的是,20 世纪 80 年代早期进行的神经心理学实验(Ullman[1981],Marr[1982])证明人类视觉上的某些方面可以以式(10.1.17)的基本形式用数学的方法进行模型化。

例 10.5 通过零交叉寻找边缘

图 10.15(a)显示了 1.3.2 节中讨论的血管造影图像。图 10.15(b)显示了这幅图像的Sobel梯度,用于进行对比。图 10.15(c)是一个用于得到 27×27 的空间平滑处理模板的空间高斯型

函数(带有一个 5 像素的标准差)。这个模板是通过以相等的间隔对高斯型函数进行取样得到的。图 10.15(d)是用于计算式(10.1.15)的空间模板。图 10.15(e)是通过对原图使用高斯型平滑模板进行平滑处理,随后使用拉普拉斯算子模板得到的LoG图像(这幅图像是除去平滑处理模板产生的边缘效应后得到的)。如前面段落中提到的,$\nabla^2 h$ 可以通过先应用图 10.15(c)后应用图 10.15(d)计算出来。应用这一过程提供了对平滑处理函数更多的控制,且经常会产生两个模板,这两个模板与直接计算式(10.1.17)的单一复合模板比起来要小得多。复合模板通常要大得多,因为复合模板必须能合并如图 10.14(a)中所示的更多的复杂形状。

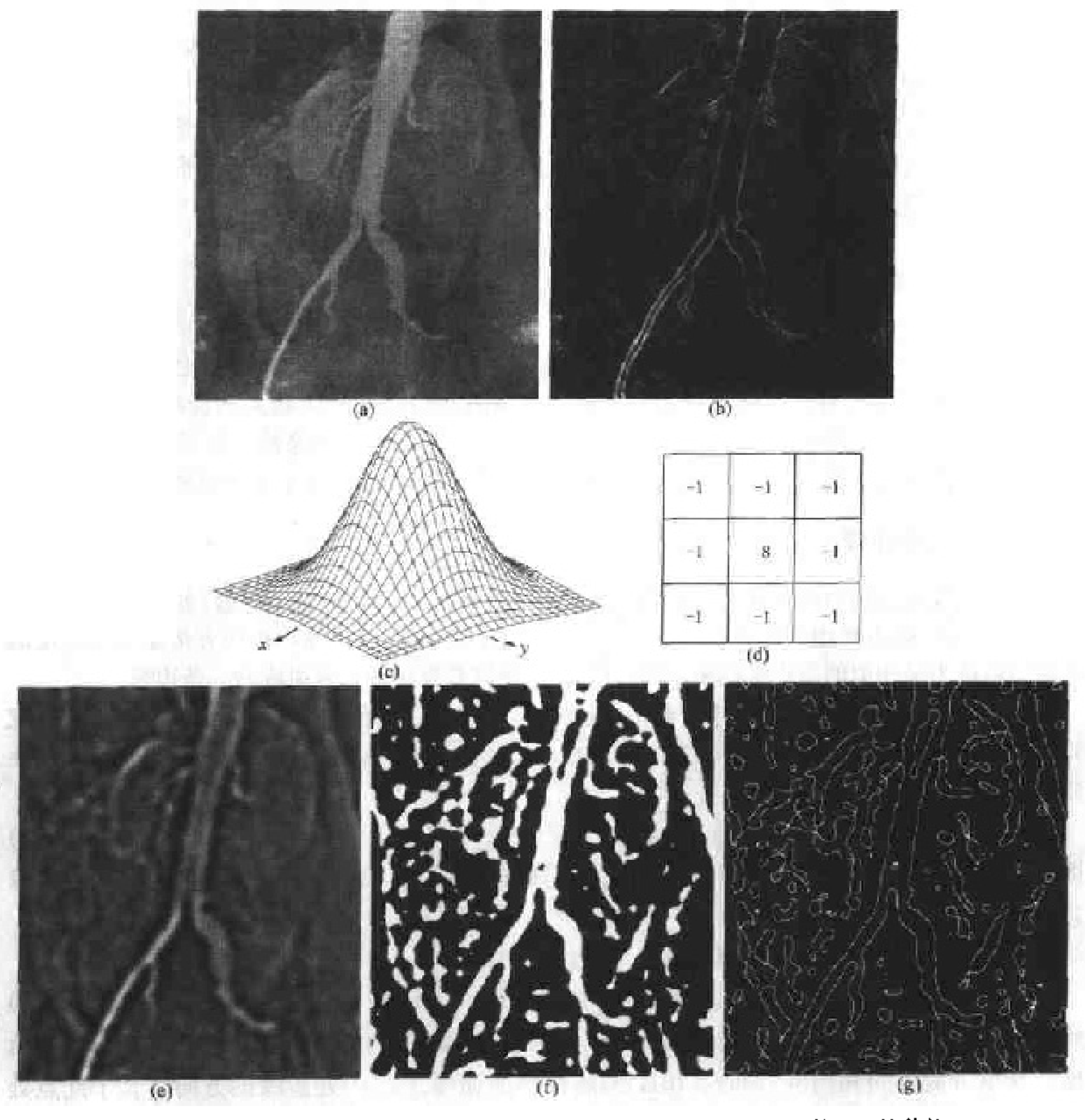


图 10.15　(a)原图,(b)Sobel 梯度(用于对比),(c)空间高斯型平滑函数,(d)拉普拉斯算子模板,(e)LoG,(f)对LoG设置门限,(g)零交点(原图由Vanderbilt大学医学中心放射学和应用辐射学系的David R.Pickens博士提供)

示于图 10.15(e)的 LoG 结果是为寻找边缘计算零交叉的一幅图像。一种对零交叉点进行近似的直接方法是通过设置LoG图像的所有正值区域为白色,负值区域为黑色来门限化。结果如图 10.15(f)所示。这种方法所隐含的逻辑是零交叉在拉普拉斯算子的正值和负值之间发生。最后,图 10.15(g)显示了估计的零交叉点,这个交叉点是通过扫描经门限处理后的图像并标记白色区域和黑色区域之间的过渡点得到的。

对比图 10.15(b)和图 10.15(g),从中可以揭示出几点有趣且重要的差异。首先,我们注意到在零交叉点图像中的边缘比梯度边缘细。这是零交叉点很引人注意的特点。另一方面,在图 10.15(g)中看到边缘是由来自许多闭合环的零交叉点所决定的。这种所谓的空心粉效应是这种方法最严重的缺陷之一。另一个主要的缺陷是零交叉点的计算。它是这种方法的基础。尽管在这个例子中这种计算相当简单,但总体上讲,这种计算相当复杂。要得到可接受的结果需要有进一步改进的技术(见Huertas和Medione[1986])。

零交叉点方法由于它抑制噪声的能力和潜在的反干扰性能而引人注目。然而,前述的局限性使其在实际应用中遇到相当大的阻碍。因此,与零交叉点方法相比,基于不同的梯度计算方法的边缘查找技术在分割算法中仍频繁应用。

10.2 边缘连接和边界检测

从理论上讲,前面一节中讨论的方法仅得到处在边缘上的像素点。实际上,由于噪声、不均匀的照明而产生的边缘间断以及其他由于引入虚假的亮度间断所带来的影响,使得到的一组像素很少能完整地描绘一条边缘。因此,典型的做法是在使用边缘检测算法后紧跟着使用连接过程将边缘像素组合成有意义的边缘。一些基本的方法很适合于这个目的。

10.2.1 局部处理

连接边缘点最简单的方法之一是分析图像中每个点(x,y)的一个小邻域(如 3×3 或 5×5)内像素的特点,该点是用前节中讨论过的某一种技术标记了的边缘点。将所有依据事先预定的准则而被认为是相似的点连接起来,形成由共同满足这些准则的像素组成的一条边缘。

在这种分析过程中确定边缘像素相似性的两个主要性质是:(1)用于生成边缘像素的梯度算子的响应强度和(2)梯度向量的方向。如式(10.1.4)和式(10.1.12)所定义的那样,第 1 条性质由∇f值给出。因此,如果满足:

$$|\nabla f(x,y)-\nabla f(x_0,y_0)|\leqslant E \tag{10.2.1}$$

则处在预先定义的(x,y)邻域内坐标为(x_0,y_0)的边缘像素,在幅度上相似于位于(x,y)的像素,这里 E 是一个非负门限。

梯度向量的方向(角度)由式(10.1.5)给出。如果:

$$|\alpha(x,y)-\alpha(x_0,y_0)|<A \tag{10.2.2}$$

则处在预先定义的(x,y)邻域内坐标为(x_0,y_0)的边缘像素具有相似于位于(x,y)的像素的角度。这里 A 是非负角门限。如式(10.1.5)所注释的那样,(x,y)处边缘的方向垂直于此点处梯度向量的方向。

如果大小和方向准则得到满足,则在前面定义的(x,y)邻域中的点就与位于(x,y)的像素连接起来。在图像中的每个位置重复这一操作。当邻域的中心从一个像素转移到另一个像素

时，这两个相连接点必须记录下来。

例10.6 基于局部处理的边缘点连接

为了说明前述过程，考虑图10.16(a)，图中显示了一辆汽车尾部的图像。目的是找到适合牌照大小的矩形。通过检测清晰的水平和垂直边缘可以构成这些矩形。图10.16(b)和(c)显示了通过使用水平和垂直Sobel算子得到的相应边缘。图10.16(d)显示了连接同时具有大于25的梯度值且梯度方向差不超过15°的所有点的结果。通过对图10.16(c)的每一行顺序地应用这些准则而得到水平方向的线。对图10.16(b)逐列扫描生成垂直方向的线。进一步的处理包括将之间由小断裂分开的边缘线段连接起来，并清除孤立的短线段。正如图10.16(d)所示，对应于牌照大小的矩形是图像中检测到的少数几个矩形之一。使用这些矩形对牌照进行定位是一个简单的问题(对于美国牌照，其长宽比例具有2:1的可识别比例)。

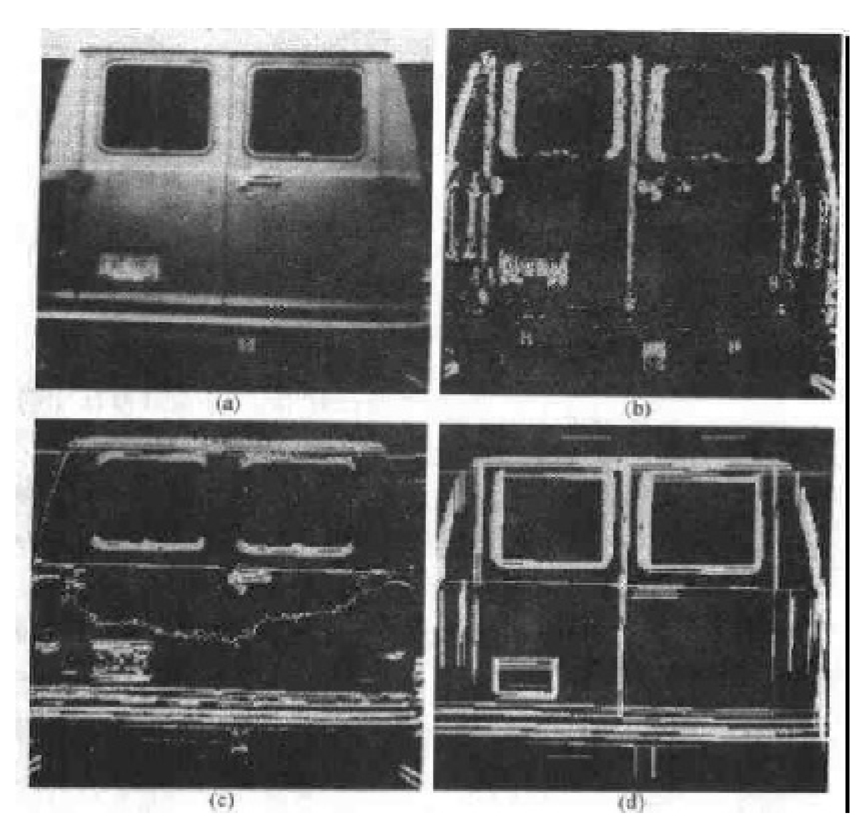


图10.16 (a)输入图像，(b)梯度的G_y分量，(c)梯度的G_x分量，(d)边缘连接的结果(由Perceptics公司提供)

10.2.2 通过霍夫变换进行整体处理

在这一节中，如果点在一条特定形状的曲线上，则先确定再进行连接。与10.2.1节中讨论的局部分析方法不同，我们现在考虑像素之间的整体关系。

在图像中给出n个点，假设我们希望找到这些点中位于直线上的点组成的子集。一种可行的方法就是先寻找所有由每对点确定的直线，然后找到所有接近特定直线的点组成的子集。这个过程的问题是它涉及寻找$n(n-1)/2 \sim n^2$条直线，并且对每个点要与所有直线执行

$(n)(n(n-1))/2 \sim n^3$次比较。这种方法在计算上是不可行的并且没有什么应用价值。

霍夫[Hough,1962]提出一种替代方法,一般称为霍夫变换。考虑一个点(x_i,y_i)和一条直线的斜截式方程 $y_i = ax_i + b$。通过点(x_i,y_i)的直线有无数条,且对不同的 a 和 b 值,它们都满足这个等式 $y_i = ax_i + b$。然而,将等式写成 $b = -x_i a + y_i$ 的形式并参考 ab 平面(也叫参数平面)将得到对于定点(x_i,y_i)的惟一直线方程。再有,在参数空间中,第 2 个点(x_j,y_j)也有与之相关的一条直线,且这条直线与(x_i,y_i)相关的直线相交于(a',b')点。这里 a'是斜率,b'是 xy 平面上包含点(x_i,y_i)和(x_j,y_j)的直线的截距。实际上,在参数空间中这条直线所包含的所有点都有过点(a',b')的线。图 10.17 说明了这些概念。

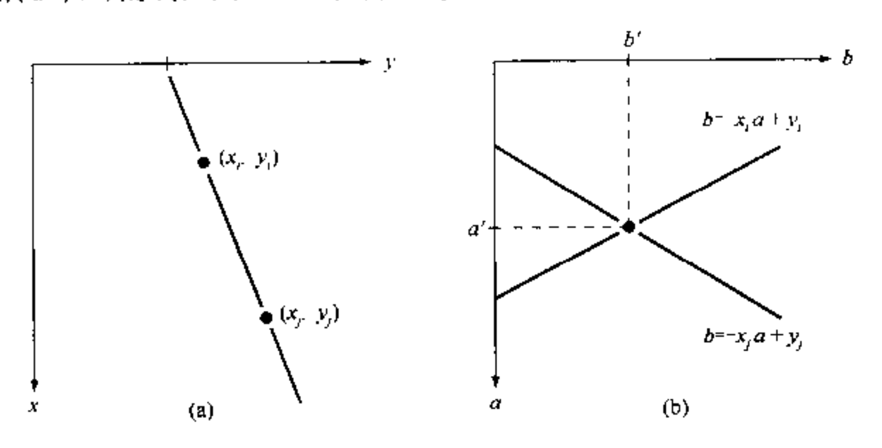


图 10.17 (a)xy 平面,(b)参数空间

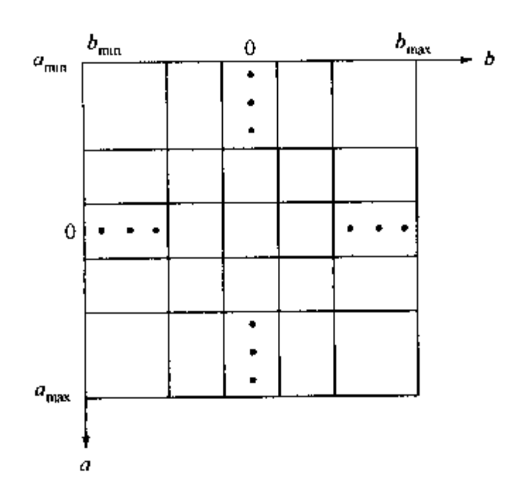


图 10.18 用于霍夫变换的参数平面的进一步分割

霍夫变换在计算上的吸引力在于将参数空间进一步分割为所谓的累加器单元,如图 10.18 所示。这里(a_{max},a_{min})和(b_{max},b_{min})分别为斜率和截距值期望的范围。位于坐标(i,j)的单元具有累加值 $A(i,j)$,并对应于与参数空间坐标(a_i,b_j)相关的矩形。开始的时候,这些单元被置为零。然后,对图像平面中的每个点(x_k,y_k),令参数 a 分别等于 a 轴上每个允许的细分值,同时使用等式得到对应的 b。然后对得到的 b 进行舍入得到 b 轴上允许的最近似的值。如果一个 a_p 值得到解 b_q,就令 $A(p,q)=A(p,q)+1$。在这一过程的最后,$A(i,j)$中的值 Q 将对应 xy 平面中直线 $y = a_i x + b_j$ 上的点 Q。在 ab 平面中细分的数目决定了这些点共线性的精确度。

注意,以 K 为增量对 a 轴进行细分,对所有点(x_k,y_k),有 K 个 b 值对应 K 个可能的 a 值。由于有 n 个图像点,所以这种方法需要 nK 次计算。因此,刚刚讨论的过程是关于 n 的线性过程,并且除非 K 接近或超过 n,否则乘积 nK 不会达到本节开始时讨论的方法的计算量。

使用等式 $y = ax + b$ 表示一条直线带来的一个问题是,当直线接近垂直时,直线的斜率接近无限大。解决这一难点的一种方法是使用直线的标准式:

$$x\cos\theta + y\sin\theta = \rho \tag{10.2.3}$$

图 10.19(a)说明了用于式(10.2.3)的参数的几何解释。使用这一表达式构造一个累加器的表与使用讨论过的斜截式方法一样。然而，与直线不同，这个公式的轨迹是 $\rho\theta$ 平面上的正弦曲线。与以前一样，在 $x\cos\theta_j + y\sin\theta_j = \rho_i$ 上的共线点集 Q 生成参量空间中交于点 (ρ_i, θ_j) 的 Q 的正弦曲线。θ 的增加和对应的 ρ 的求解，给出了由 (ρ_i, θ_j) 决定的单元相关累加器 $A(i, j)$ 中 Q 的输入。图 10.19(b)说明了参数空间的划分。

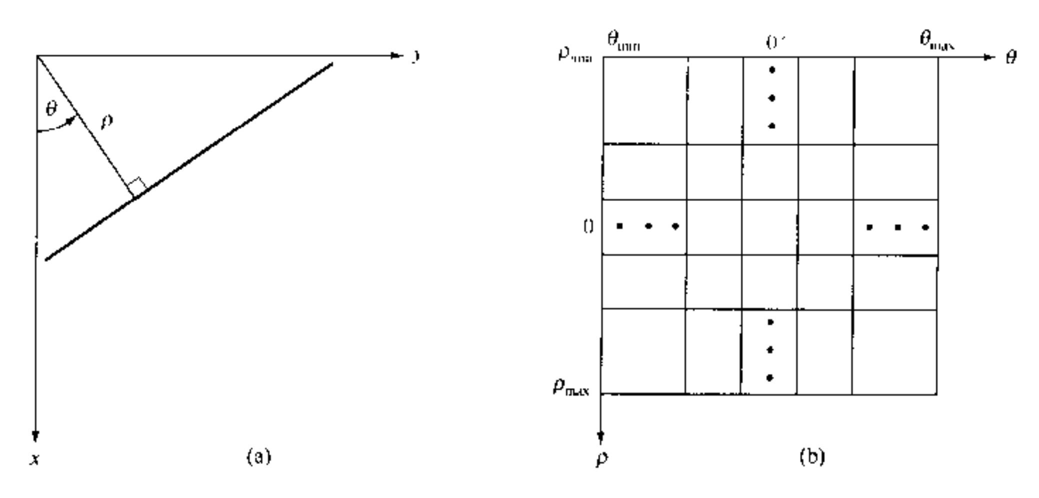


图 10.19　(a)直线的标准表达方法，(b)将 $\rho\theta$ 平面细分为不同单元

角 θ 的取值范围为 $\pm 90°$，以 x 轴为基准。因此，参考图 10.19(a)，水平线的角度 $\theta = 0°$，ρ 等于正的 x 截距。同样，垂直线的角度为 $\theta = 90°$，ρ 等于正的 y 截距，或 $\theta = -90°$，ρ 等于负的 y 截距。

例 10.7　霍夫变换的说明

图 10.20 说明了霍夫变换是基于式(10.2.3)的。图 10.20(a)显示了一幅带有 5 个标记点的图像。每个点都映射到 $\rho\theta$ 平面上，如图 10.20(b)所示。角 θ 的取值范围为 $\pm 90°$，且 ρ 轴的范围为 $\pm\sqrt{2}D$，这里 D 是图像对角之间的距离。与基于使用斜率截距的变换不同，这里每条曲线都有一个不同的正弦形状。从 1 值点映射得到的水平线是具有零幅值的正弦的特殊情况。

霍夫变换的共线性检测性质在图 10.20(c)中给予说明。点 A(不要同累加器值相混淆)代表 xy 图像平面内对应于点 1,3 和 5 的曲线的交点。点 A 的位置表示这 3 个点在一条过原点($\rho = 0$)且方向为 $-45°$的直线上。同样，在参数空间中相交于点 B 的曲线表示点 2,3 和 4 位于方向为 45°且与原点距离为 1/2 对角线长度(即从图像的原点到相对的角)的直线上。

最后，图 10.20(d)说明了这样的事实，即霍夫变换展示了参数空间左右边缘的反射式的邻接关系。在图 10.20(d)中由标记为 A，B 和 C 的点说明的这种性质是 θ 和 ρ 在 $\pm 90°$边界附近变号引起的结果。

尽管到目前为止我们关注的焦点始终在直线上，但霍夫变换也适用于任何形式为 $g(\mathbf{v}, \mathbf{c}) = 0$的函数，这里 $\mathbf{v}$ 是坐标向量，$\mathbf{c}$ 是系数向量。例如，位于圆上的点：

$$(x - c_1)^2 + (y - c_2)^2 = c_3^2 \tag{10.2.4}$$

可以通过使用刚才讨论的方法进行检测。主要的差别是存在3个参数(c_1,c_2 和 c_3),这3个参数生成了一个带有立方单元和形如 $A(i,j,k)$ 的累加器的三维参数空间。这个过程是增加 c_1 和 c_2,解出满足式(10.2.4)的 c_3,并更新对应于三元组(c_1,c_2,c_3)相关单元的累加器。很清楚,霍夫变换的复杂性是与给定函数表达式的坐标和系数的数目成比例的。霍夫变换的进一步推广使我们可以不用简单分析表达式就可以对曲线进行检测,正如灰度级图像的变换的应用。处理这些扩展问题的参考资料列于本章末尾部分。

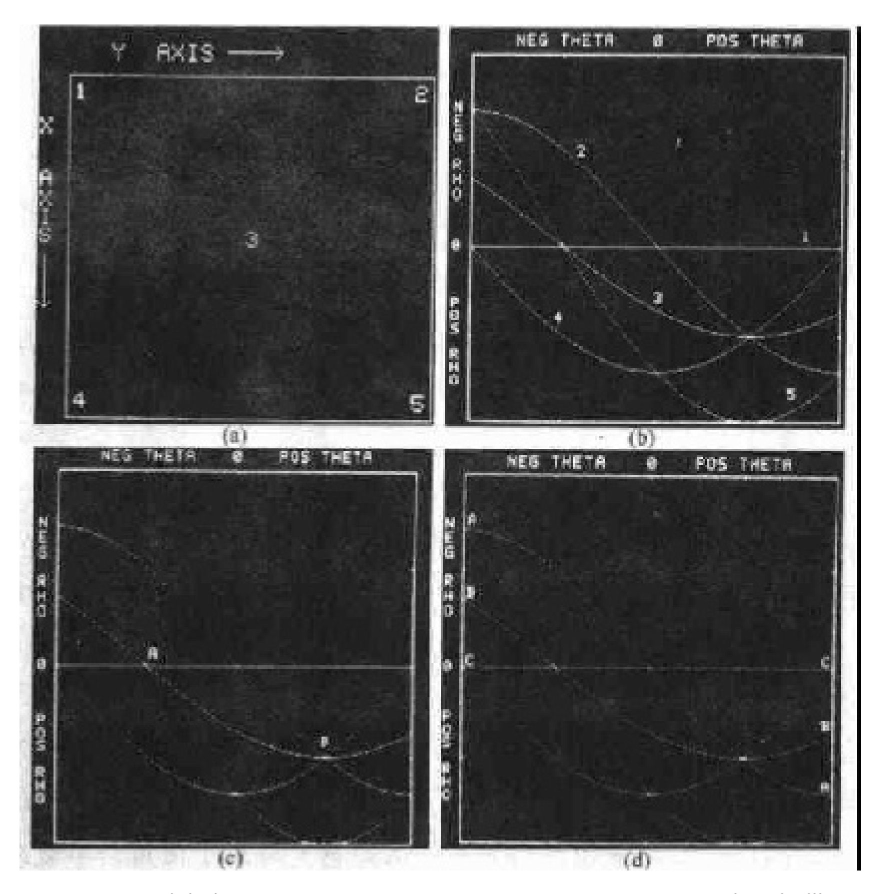


图 10.20 霍夫变换的说明(Texas Instruments 公司的 D.R.Cate 先生提供)

我们现在回到边缘连接问题上来。一种基于霍夫变换的连接方法如下所述:

1. 计算图像的梯度并对其设置门限得到一幅二值图像。
2. 在 $\rho\theta$ 平面内确定再细分。
3. 对像素高度集中的地方检验其累加器单元的数目。
4. 检验选择的单元中像素之间的关系(主要针对连续性)。

此时连续性的概念通常基于不连续像素间距离的计算,这些不连续像素是在遍历给定累加器单元对应的像素集合的过程中识别出来的。如果一点和它最接近的相邻点之间的距离超过某一门限,就认为点之间的间隔是有意义的(见2.5节中连通性、邻域和距离测量的讨论)。

例 10.8 使用霍夫变换进行边缘连接

图10.21(a)显示了一幅航拍的红外线图像,图像中包含两个机库和一条跑道。

图 10.21(b)是设置了门限的梯度图像,此图像是使用了10.1.3节中讨论的Sobel算子得到的(注意在跑道边缘的细小的缝隙)。图 10.21(c)显示了梯度图像的霍夫变换。图 10.21(d)显示了(在白色区域)依据下列准则判断为相连的像素集合:(1)这些像素属于3个具有最高计数的累加器单元之一和(2)没有长于5个像素的间隙。注意没有间隙的出现是连接的结果。

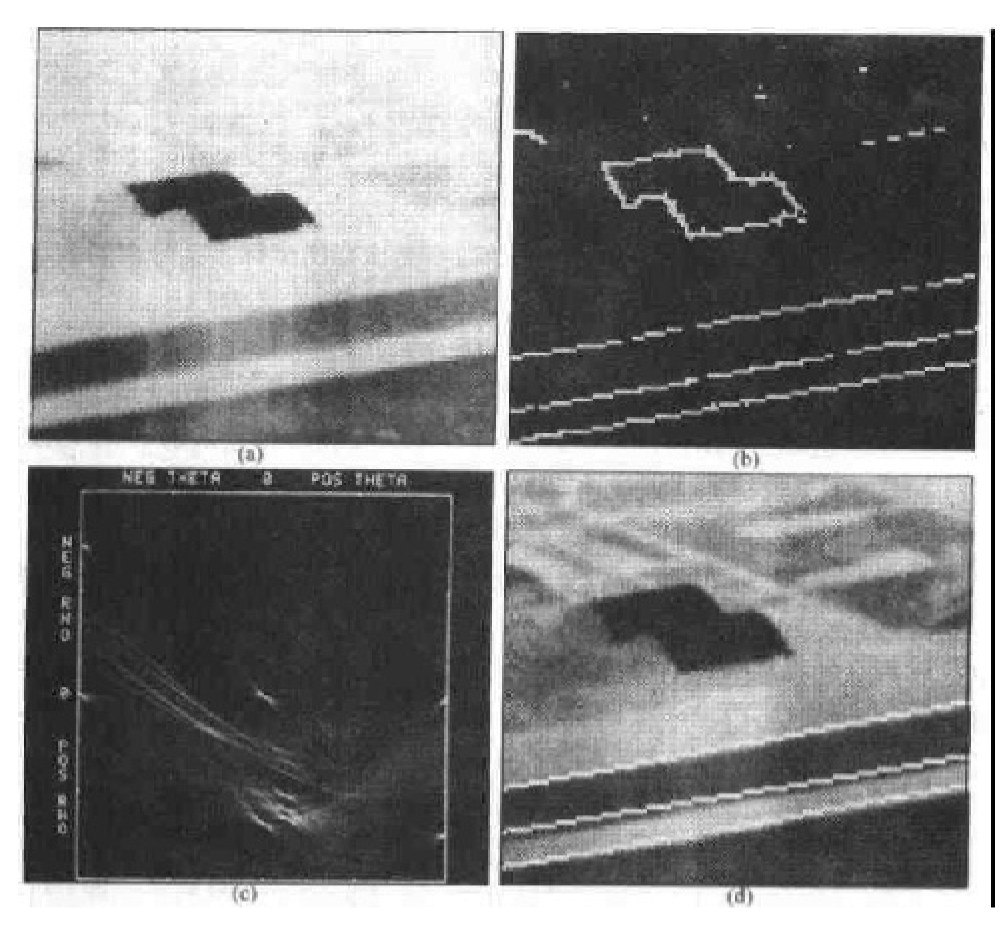


图 10.21 (a)红外线图像,(b)设置了门限的梯度图像,(c)霍夫变换,(d)连接的像素(Texas Instruments 公司的 D.R.Cate 先生提供)

10.2.3 通过图论技术进行全局处理

在本节中,我们讨论一种全局性的边缘检测方法和基于用图形方式表达边缘线段的连接,以及搜索与重要边缘相对应的低开销路径的图形。这种表示提供了一种在有噪声环境下效能很好的抗干扰途径。如所期望的那样,这个过程更为复杂并且是迄今为止已讨论过的方法中处理时间消耗最大的方法。

我们从一些基本的定义开始展开。图 $G=(N,U)$是一个有限非空节点集合 N 与一个同 N 性质不同的无序点对集合 U。U 中的每对(n_i,n_j)称为一条弧。被检测到弧的图称为指向图。如果一条弧从节点 n_i 指向 n_j,则称 n_j 为父亲节点 n_i 的后继节点。识别一个节点的后继节点称为节点的扩展。在每幅图中对节点定义不同的级别,比如第 0 级包括一个单节点,称为开始或根节点,最后一级节点称为目的点。开销 $c(n_i,n_j)$可以同弧(n_i,n_j)相联系。节点序列

$n_1, n_2, \cdots, n_k$ 称为从 n_1 到 n_k 的路径,其中每个节点 n_i 是节点 n_{i-1} 的后继节点。整条路径的开销为:

$$c = \sum_{i=2}^{k} c(n_{i-1}, n_i) \tag{10.2.5}$$

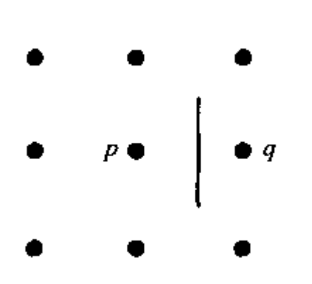


图 10.22 像素 p 和 q 之间的边缘元素

如果将边缘元素定义为像素 p 和 q 之间的边界,p 和 q 就是如图 10.22 所示的四邻接像素的点。这样下面的讨论就简单了。边缘元素由点 p 和 q 的 xy 坐标来识别。换言之,图 10.22 中的边缘元素由点对 $(x_p, y_p)(x_q, y_q)$ 定义。为了与 10.1.3 节中给出的定义保持一致,边缘定义为相连的边缘元素序列。

我们可以用图 10.23(a)中显示的 3×3 图像说明刚才提到的这些概念是如何用于边缘检测的。边框外部的数字是像素的坐标。括号内的数字表示灰度值。每条由像素 p 和 q 定义的边缘元素都有一个相应的开销,定义为:

$$c(p, q) = H - [f(p) - f(q)] \tag{10.2.6}$$

这里 H 是图像中最高的灰度值(此图中为 7),$f(p)$ 和 $f(q)$ 分别是 p 和 q 的灰度值。按惯例,点 p 位于沿边缘元素追踪方向的右手一边。例如,边缘线段(1,2)(2,2)在图 10.23(b)中点(1,2)和(2,2)之间。如果追踪方向是向右的,则点 p 坐标为(2,2)而点 q 坐标为(1,2),所以 $c(p, q) = 7 - [7 - 6] = 6$。这个开销显示在边缘线段下方的方框中。另外,如果我们在两个相同的点之间向左追踪,则点 p 坐标为(1,2),而点 q 坐标为(2,2)。在这种情况下,开销为 8,如图 10.23(b)边缘线段上方所示的那样。为简化讨论,假设边缘从最顶部的一行开始并结束于最后一行。这样边缘的第 1 个元素只能在点(1,1),(1,2)或(1,2),(1,3)之间。同样,最后一个边缘元素只能在点(3,1),(3,2)或(3,2),(3,3)之间。记住 p 和 q 如前所述是四邻接的。

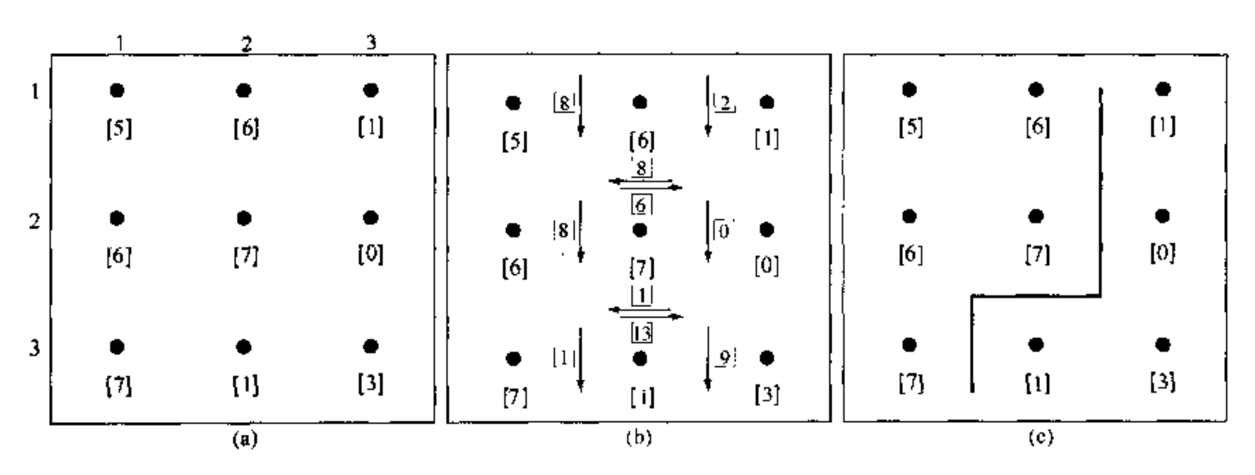


图 10.23 (a)一个 3×3 图像区域,(b)边缘线段和相应的开销,(c)对应于图10.24中显示的最小开销路径的边缘

图 10.24 显示了反映这个问题的图。图中每个节点(矩形)对应于图 10.23 中的一个边缘元素。如果前后紧随的两条相应的边缘元素是同一条边缘的一部分,则两个节点之间存在一条弧。

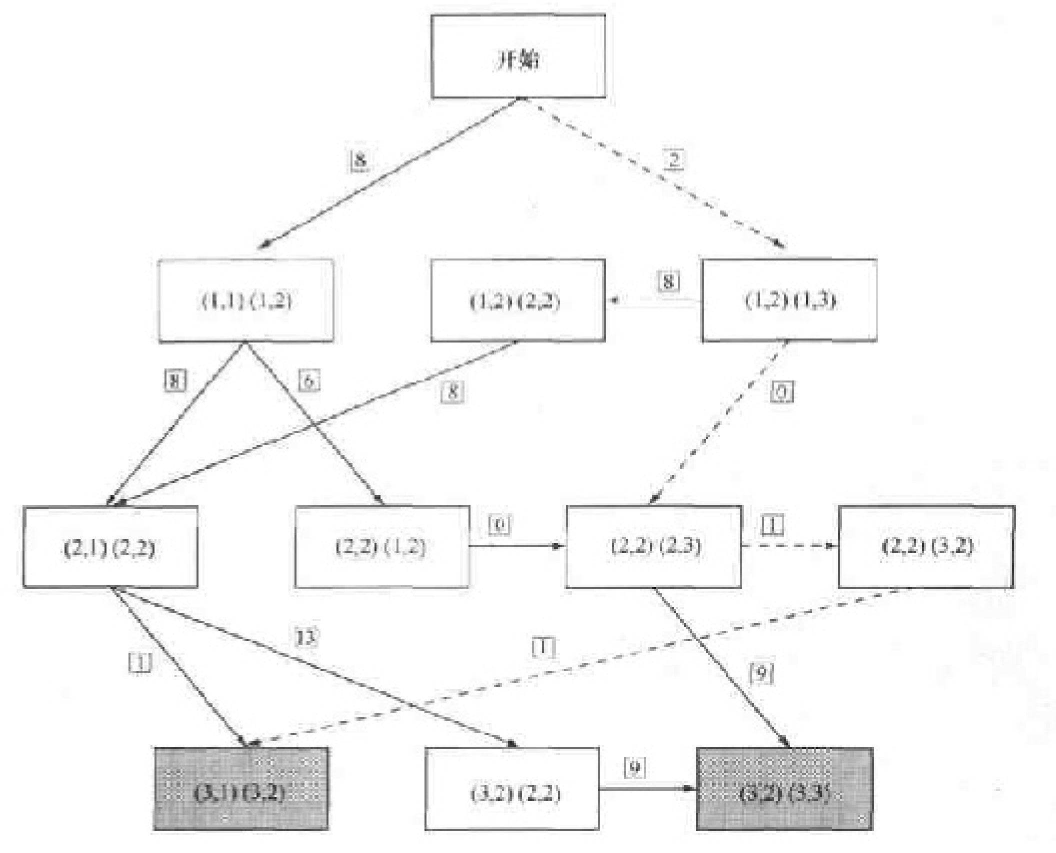


图 10.24　图 10.23(a)中图像的图，虚线表示最小开销路径

如图 10.23(b)所示，使用式(10.2.6)计算出的每条边缘线段的开销示于指向对应节点的弧旁边的方框中。目的节点以暗色表示。最小开销路径用虚线表示，图 10.23(c)显示了对应于这条路径的边缘。

一般来讲，寻找最小开销路径问题就计算来看是很重要的。这种方法是典型的为速度牺牲最优性的方法。下面的算法代表了一类方法。这种方法为了减少搜寻的工作量而使用试探性的方法进行。令 $r(n)$为从起点 s 到一个目标节点的最小开销路径的开销估计。这里路径必须通过 n。这样路径的开销可以表示为，从起点 s 到 n 的最小开销路径的开销估计，加上从 n 到一个目标节点的路径的开销估计。即：

$$r(n) = g(n) + h(n) \tag{10.2.7}$$

这里，$g(n)$可以作为目前找到的从 s 到 n 的最小开销路径的选择，通过使用任何可用的试探信息(比如基于得到节点的前一个开销只扩展一定的节点)而得到 $h(n)$。一种以 $r(n)$为基础进行图搜索的算法如下：

步骤 1：标记开始节点为“开(OPEN)”并置 $g(s)=0$。

步骤 2：如果不存在标记为“开”的节点，则失败；否则继续。

步骤 3：标记最接近“开”节点的节点 n 为“关(CLOSED)”，用式(10.2.7)计算得出的此节点的估计值 $r(n)$最小(就最小值 r 进行的连接可以采用任何方法进行，但必须总是有利于到达目标节点)。

步骤 4：如果 n 是一个目标节点，通过使用指针进行反向追踪得到解决路径则退出；否则继续。

步骤 5:扩展节点 n,得到它的所有后继节点(如果没有后继节点则转到步骤 2)。

步骤 6:如果一个后继节点 n_i 未被标记,则设置:

$$r(n_i) = g(n) + c(n, n_i)$$

并将它标记为“开”,并直接将指针反向指向 n。

步骤 7:如果一个后继节点 n_i 被标记为“关”或“开”,通过下式更新此节点的值:

$$g'(n_i) = \min[g(n_i), g(n) + c(n, n_i)]$$

将这些标记为“关”的后继节点标记为“开”,其 g' 值被降低,并且从所有那些 g' 值被降低的节点指针重新指向 n。转到步骤 2。

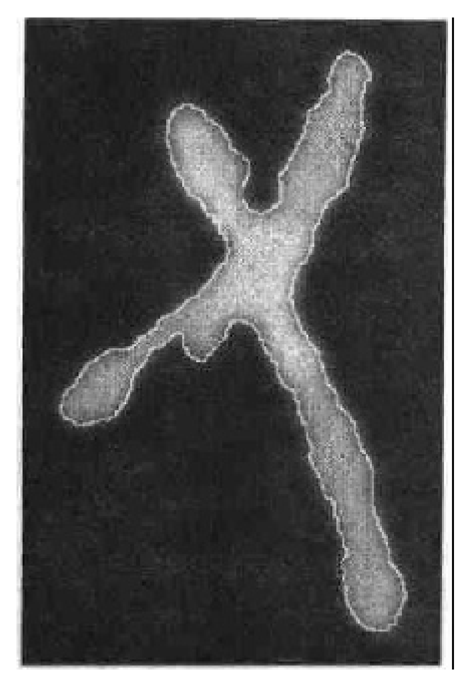


图 10.25 带干扰的染色体轮廓图和由图搜索决定的边缘边界(用白色表示)

这个算法并不能保证得到一个最小开销路径;它的优点在于通过试探而得到速度优势。然而,如果 $h(n)$ 对于从节点 n 到一个目标节点的最小开销路径的开销来说是一个更低的界限,这个方法确实能得到一个通向目标的最佳路径(Hart等[1968])。如果没有可用的试探信息(即 $h\equiv 0$),则这个过程退化成Dijkstra[1959]提出的均匀开销算法。

例 10.9 通过图搜索寻找边缘

图 10.25 显示了一幅带干扰的染色体轮廓图和一条使用试探性的图搜索方法找到的边缘。搜索方法是基于本节中探讨的算法。边缘用白色表示,并重叠于原图像上。注意,此时对象的边缘和边界几乎是相同的。开销用式(10.2.6)算出,并且对图像中每个点的试探,确定并使用了从被试探点开始向下经过的包含 5 个层次的最佳路径。考虑到图像中存在噪声量,图搜索方法生成了一个可接受的相当准确的结果。

10.3 门限处理

由于图像门限处理的直观性和易于实现的性质,使它在图像分割应用中处在中心地位。在 3.1 节中我们第一次介绍了简单的门限处理。在前面的章节中,我们将这种处理应用于对不同方面的讨论中。本节中,我们将以更为正式的方式介绍门限处理,并将这种处理技术进行推广,使其相对于以前所论述的内容来说更为一般化。

10.3.1 基础

假设图 10.26(a)所示的灰度级直方图对应于一幅图像 $f(x,y)$。这幅图像由亮的对象和深的背景组成,这样的组成方式将对象和背景具有灰度级的像素分成两组不同的支配模式。从背景中提取对象的一种显然的方法是选择一个门限值 T,将这些模式分离开。然后,所有 $f(x,y)>T$的点(x,y)称为对象点;否则,就称为背景点。这是 3.1 节中介绍的门限处理类型。

图 10.26(b)显示了这种方法更为一般化的情况。这里3个主模式描绘了图像的直方图特性(例如,在暗色背景上的两类亮色对象)。这里,多门限处理把一个点分类,如 $T_1 < (x,y) \leq T_2$ 时将点分为某一对象,而如果 $f(x,y) > T_2$ 时则归为另一个对象,如果 $f(x,y) \leq T_1$ 时归为背景。总的来说,需要多个门限的分割问题用区域生长方法能得到最好的解决。比如10.4节中讨论过的那些方法。

基于前面的论述,门限处理也许被看做一种涉及测试下列形式函数 T 的一种操作。

$$T = T[x,y,p(x,y),f(x,y)] \tag{10.3.1}$$

这里 $f(x,y)$ 是点 (x,y) 的灰度级,$p(x,y)$ 表示这个点的局部性质——例如,以 (x,y) 为中心的邻域的平均灰度级。经门限处理后的图像 $g(x,y)$ 定义为:

$$g(x,y) = \begin{cases} 1 & f(x,y) > T \\ 0 & f(x,y) \leq T \end{cases} \tag{10.3.2}$$

因此,标记为1(或其他任何合适的灰度级)的像素对应于对象,而标记为0(或任何其他没有被标记为对象的灰度)的像素对应于背景。

当 T 仅取决于 $f(x,y)$(即,仅取决于灰度级值)时,门限就称为全局的。如果 T 取决于 $f(x,y)$ 和 $p(x,y)$,门限就是局部的。另外,如果 T 取决于空间坐标 x 和 y,门限就是动态的或自适应的。

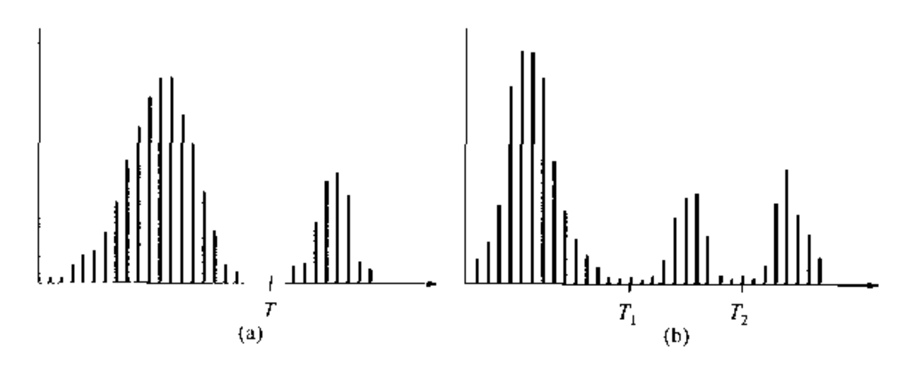


图 10.26 可以用(a)单一的门限和(b)多门限进行分割的灰度级直方图

10.3.2 亮度的作用

在2.3.4节中,我们介绍了一种简单的模型,在这种模型中,一幅图像 $f(x,y)$ 是由反射率分量 $r(x,y)$ 和亮度分量 $i(x,y)$ 的乘积组成的。本节的目的是使用这种模型简要讨论亮度对门限处理的影响,特别是对全局门限处理的影响。

考虑图 10.27(a)中计算机生成的反射函数。图 10.27(b)是这个函数的直方图,这个图很明显是双峰的并且可以通过在直方图波谷位置设置单一的全局门限 T 进行分割。用图 10.27(c)中的亮度函数与图 10.27(a)中的反射函数相乘,得到图 10.27(d)中的图像。图 10.27(e)显示了这幅图像的直方图。注意原来的波谷实际上被削去了。这使图像不可能通过单一门限进行分割。尽管我们很少使用反射函数本身对图像进行处理,但这个简单的例子说明对象和背景的反射性质可以使它们易于分离。然而,亮度条件不好的图像(如本例中的不均匀的情况)是难于分割的。

图 10.27(e)中的直方图有很大失真的原因在参考4.5节的讨论的基础上可以得到解释。

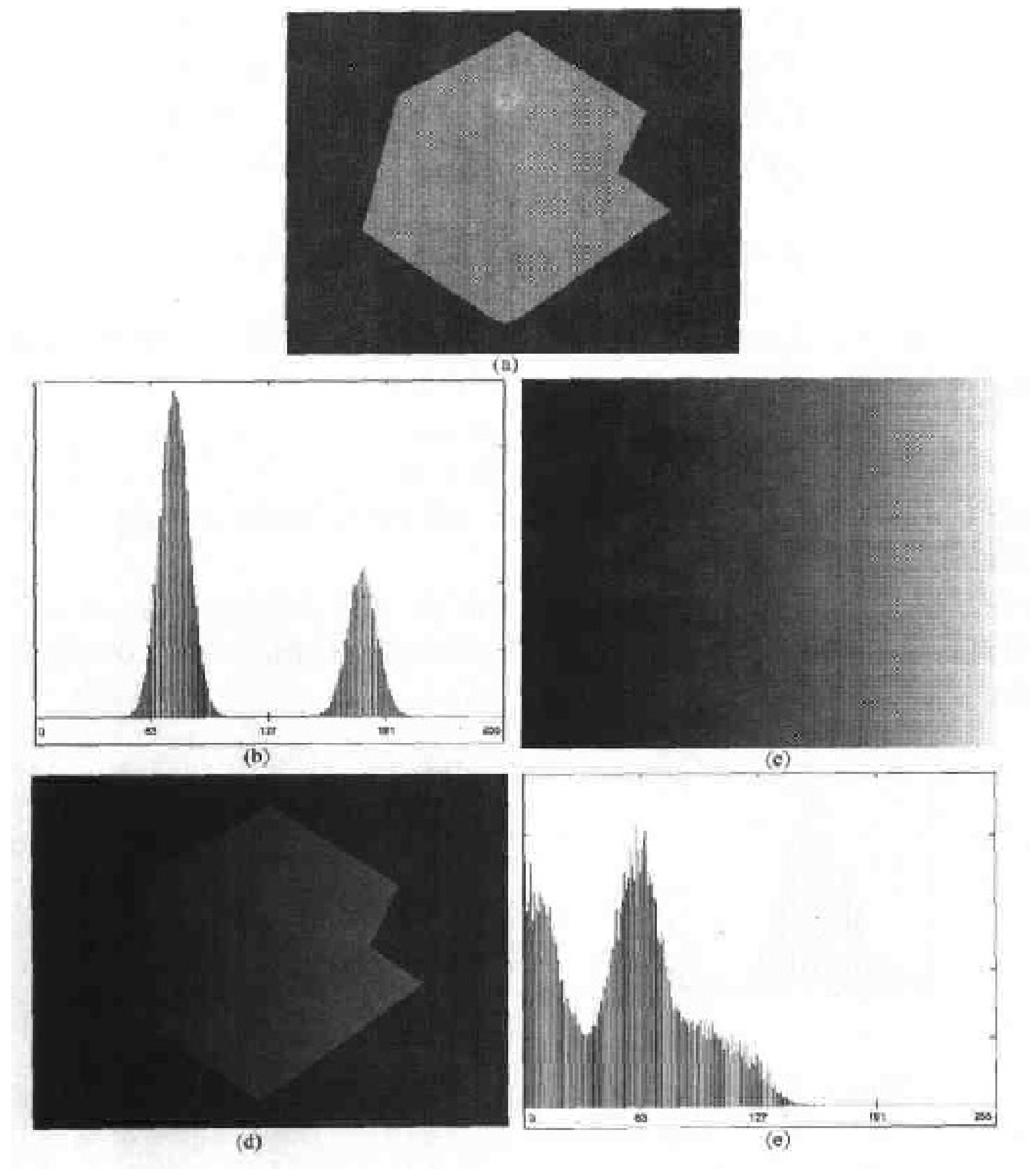


图 10.27 (a)计算机生成的反射函数,(b)反射函数的直方图,(c)计算机生成的亮度函数,(d)(a)和(c)的乘积,(e)乘积图像的直方图

由式(4.5.1),

$$f(x,y) = i(x,y)r(x,y) \tag{10.3.3}$$

取这个公式的自然对数得到和:

$$\begin{aligned} z(x,y) &= \ln f(x,y) \\ &= \ln i(x,y) + \ln r(x,y) \\ &= i'(x,y) + r'(x,y) \end{aligned} \tag{10.3.4}$$

根据概率理论(Papoulis[1991]),如果 $i'(x,y)$和 $r'(x,y)$是独立的随机变量,则 $z(x,y)$的直方图就依据 $i'(x,y)$和 $r'(x,y)$的直方图卷积给出。如果 $i(x,y)$是常量,则 $i'(x,y)$也是常量,且它的直方图是一个简单的尖峰(类似脉冲信号)。这个类似脉冲函数与 $r'(x,y)$直方图的卷积将保留这一不变的直方图基本形状(回顾4.2.4节中提到的,一个函数与一个脉冲卷积在脉

冲的位置上复制了该函数)。但如果 $i'(x,y)$ 有稍宽的直方图(不均匀亮度产生的结果),卷积处理会抹去 $r'(x,y)$ 的直方图,生成 $z(x,y)$ 的直方图,这个直方图的形状与 $r'(x,y)$ 的直方图形状有很大不同。失真的程度取决于 $i'(x,y)$ 直方图的宽度,而 $i'(x,y)$ 直方图的宽度又取决于亮度函数的不均匀性。

我们已经处理了 $f(x,y)$ 的对数,而不是直接处理图像函数,但这个问题的要点是通过使用取对数方法分离亮度分量和反射分量得到清晰解释的。这种方法允许将直方图形式看做一个卷积过程,因此可以解释,为什么反射函数的直方图中,可区分的波谷会由异常的亮度而被抹去。

当可得到的光源是可用的时候,实际中经常用的一种用于补偿不均匀性的解决方法是,将亮度图投影到一个不变的、白色反射面上。这样做生成一幅图像 $g(x,y)=ki(x,y)$,这里 k 是取决于反射面的常量,而 $i(x,y)$ 是亮度图。对于得到的具有相同亮度函数的任何图像 $f(x,y)=i(x,y)r(x,y)$,简单地用 $g(x,y)$ 除 $f(x,y)$ 得到一个标准函数 $h(x,y)=f(x,y)/g(x,y)=r(x,y)/k$。这样,如果 $r(x,y)$ 能用一个单一的门限 T 进行分割,则 $h(x,y)$ 可以用一个单一的值为 T/k 的门限分割。

10.3.3 基本全局门限

参考 10.3.1 节中的讨论,所有门限处理技术中最简单的,是使用单一的全局门限分割图像的直方图,正如在图 10.26(a)中说明的那样。这样,通过对图像进行逐像素扫描并将像素标记为对象或背景就实现了对图像的分割,对像素的标记取决于像素的灰度级大于还是小于 T 值。正如前面指出的那样,这种方法能否成功完全取决于图像的直方图能否被很好地分割。

例 10.10 *全局门限处理*

图 10.28(a)显示了一幅简单的图像,图 10.28(b)显示了它的直方图。图 10.28(c)显示了使用门限 T 分割图 10.28(a)得到的结果,门限 T 是灰度级最大值和最小值的中间值。这个门限去除了阴影部分,只留下对象本身,实现了对图像“干净的”分割。此时,我们关注的对象比背景更暗,所以,任何灰度级 $\leq T$ 的像素都被标记为黑色(0),任何灰度级 $> T$ 的像素被标记为白色(255)。关键目的是生成一幅二值图像,以便倒转黑白关系。

能够预期,上述类型的全局门限处理能够在高度可控环境中得到成功应用。其可能的领域之一就是在工业检测方面的应用,在这种应用环境中,亮度控制通常是可行的。

前述例子中的门限以直方图视觉检测为基础,通过一种试探性的方法确定的。下面的算法可以用于自动地得到 T:

1. 选择一个 T 的初始估计值。
2. 用 T 分割图像。这样做会生成两组像素:G_1 由所有灰度值大于 T 的像素组成,而 G_2 由所有灰度值小于或等于 T 的像素组成。
3. 对区域 G_1 和 G_2 中的所有像素计算平均灰度值 μ_1 和 μ_2。
4. 计算新的门限值:

$$T=\frac{1}{2}(\mu_1+\mu_2)$$

5. 重复步骤 2 到 4,直到逐次迭代所得的 T 值之差小于事先定义的参数 T_o。

当有理由认为背景和对象在图像中占据的面积相近时,则好的 T 初始值就是图像的平均灰度值。如果对象所占的面积小于背景(反之亦然),则一个像素组会在直方图中占主要地位,平均灰度值就不是好的初始选择。此时,T 更合适的初值是诸如灰度值的中间值一类的初值。参数 T_o 用于在参数 T 变得很小时停止算法的运行。这个参数在迭代的速度问题比较重要的时候应用。

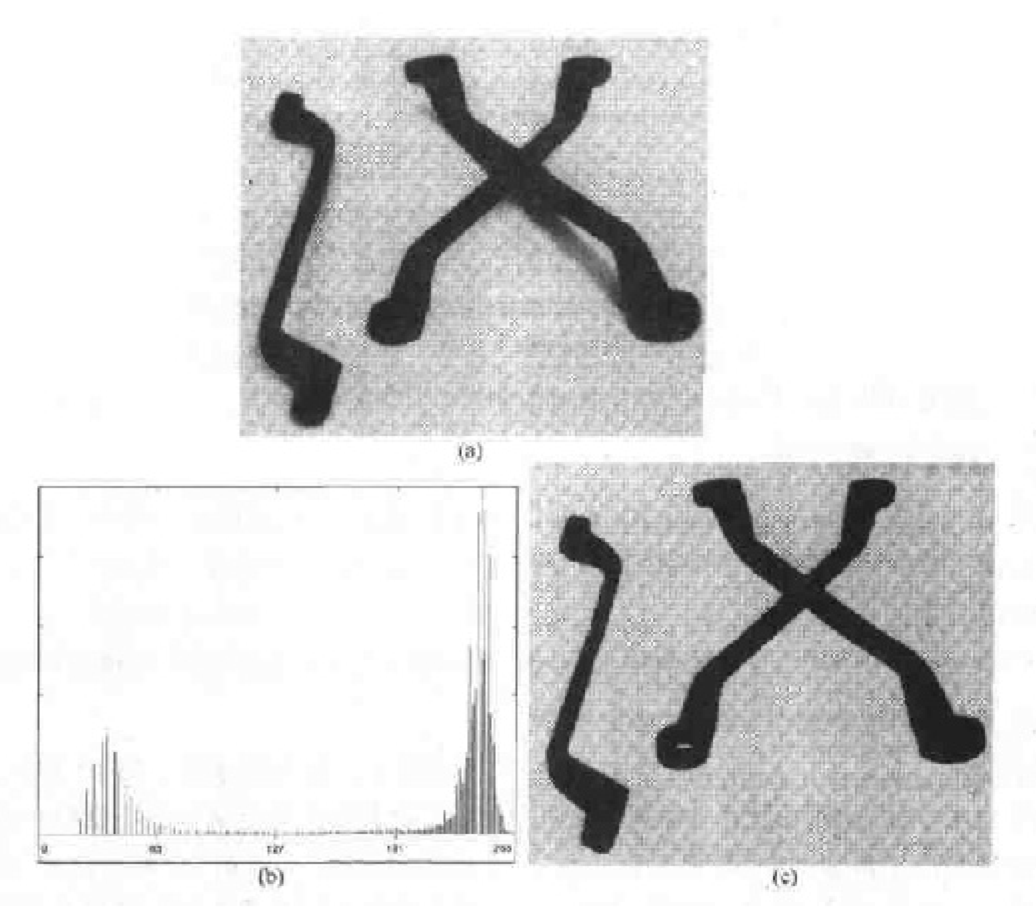


图 10.28 (a)原图,(b)图像的直方图,(c)使用全局门限 T 处理图像得到的结果,门限 T 是灰度级最大值和最小值的中间值

例 10.11 使用估计的全局门限进行图像分割

图 10.29 显示了使用基于前述算法估计的全局门限进行图像分割的例子。图 10.29(a)是原图,图 10.29(b)是图像的直方图。请注意直方图清晰的波谷。迭代算法的应用在初始时使用平均灰度级,同时令 $T_o = 0$ 的情况下,进行 3 次迭代得到值 125.4 的结果。图 10.29(c)中显示的图像是用 $T = 125$ 分割原图所得的结果。如我们从用直方图的分割方式所预料的那样,在对象和背景之间的分割是非常有效的。

10.3.4 基本自适应门限

如图 10.27 中的说明那样,像不均匀亮度这样的成像因素会导致本来可以进行很有效分割的直方图变成用单一全局门限无法有效分割的直方图。一种处理这种情况的方法就是将图像进一步细分为子图像,并对不同的子图像使用不同的门限进行分割。这种方法的关键问题

是如何将图像进行细分和如何为得到的子图像估计门限值。由于用于每个像素的门限取决于像素在子图像中的位置,这类门限处理是自适应的。我们用一个简单的例子对自适应门限处理进行说明。在下一节中我们将给出一个更综合性的例子。

例 10.12　基本的自适应门限处理

图 10.30(a)显示了来自图 10.27(d)的图像,对这幅图像我们认为无法用一个单一的全局门限进行有效的处理。实际上,图 10.30(b)显示了用一个全局门限处理后得到的结果,此门限是用人工在直方图的波谷处设置的,见图 10.27(e)。减少亮度不均匀性带来的影响的一种方法是,将图像细分为更小的子图像,使每个子图像内的亮度近似于均匀。图 10.30(c)显示了这样分割的图像。这幅图像是通过将图像细分为 4 个大小相等的部分,然后再对每个子图像重复进行四等分得到的。

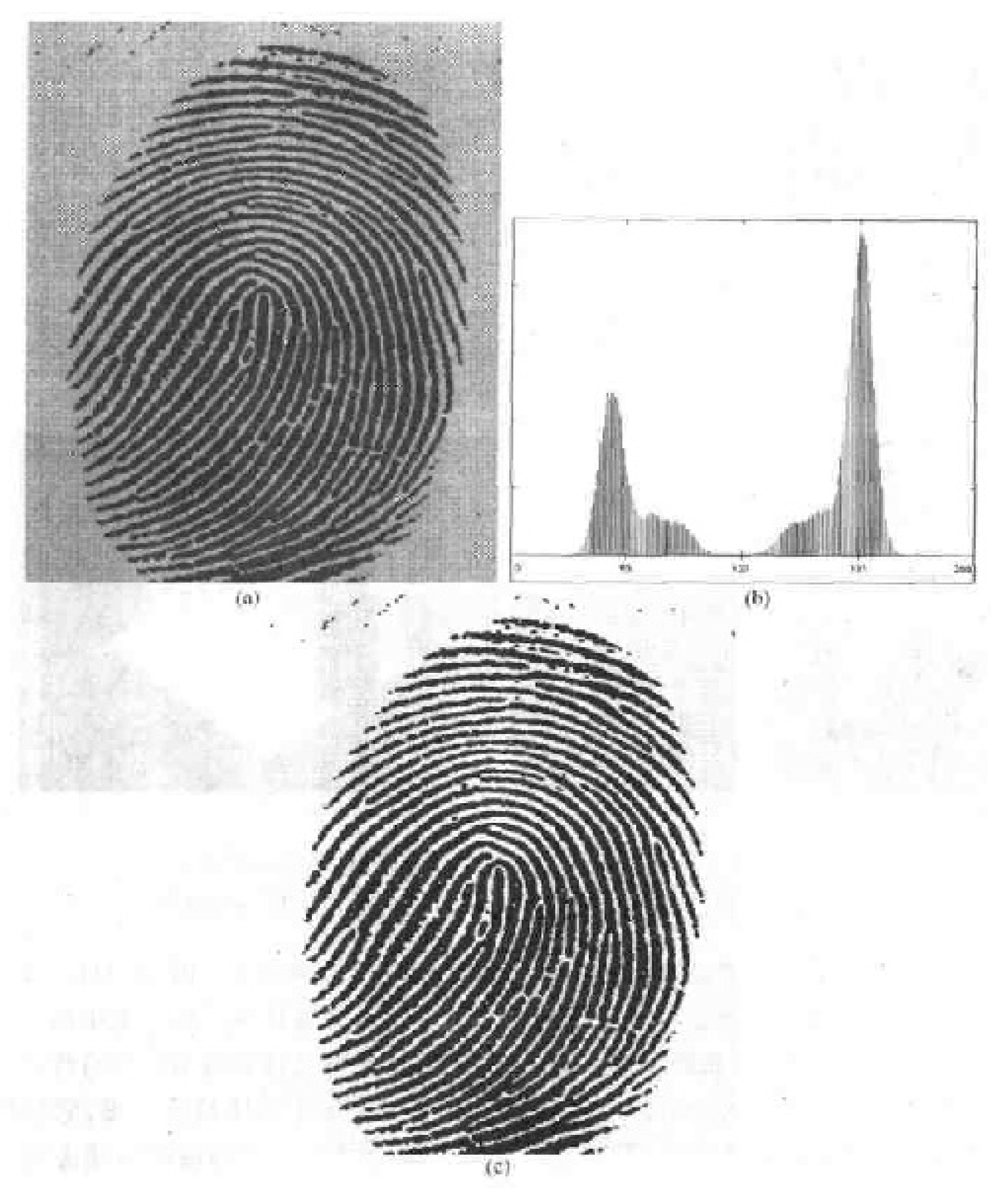


图 10.29　(a)原图,(b)图像的直方图,(c)通过用迭代估计的门限对图像进行分割的结果(由美国国家标准技术研究所提供)

所有不包含对象和背景间边界的子图像都具有小于75的方差。所有包含边界的子图像都具有大于100的方差。每幅方差大于100的子图像都用专门为本幅子图像计算的门限进行分割,门限的计算用前一节中提到的算法。处理每幅子图像时选取子图中灰度级最大值和最小值的中间值作为 T 的初值。所有方差小于100的子图像作为一幅合成图像,用与前面相同的算法估计的单一门限进行分割。

经过这一过程得到的分割结果显示于图10.30(d)中。除了两幅子图像外,对图10.30(b)的改进之处是很明显的。

分割不当的子图像中对象和背景间的边界细而暗,而且得到的直方图几乎都是单峰的。图10.31(a)显示了来自图10.30(c)顶部未经适当分割的子图像和位于此图上面经过适当分割的子图像。经过适当分割的子图像的直方图很明显是双峰的,且具有明确的波峰和波谷。另一个直方图近于单峰,在对象和背景之间没有明确的区分。

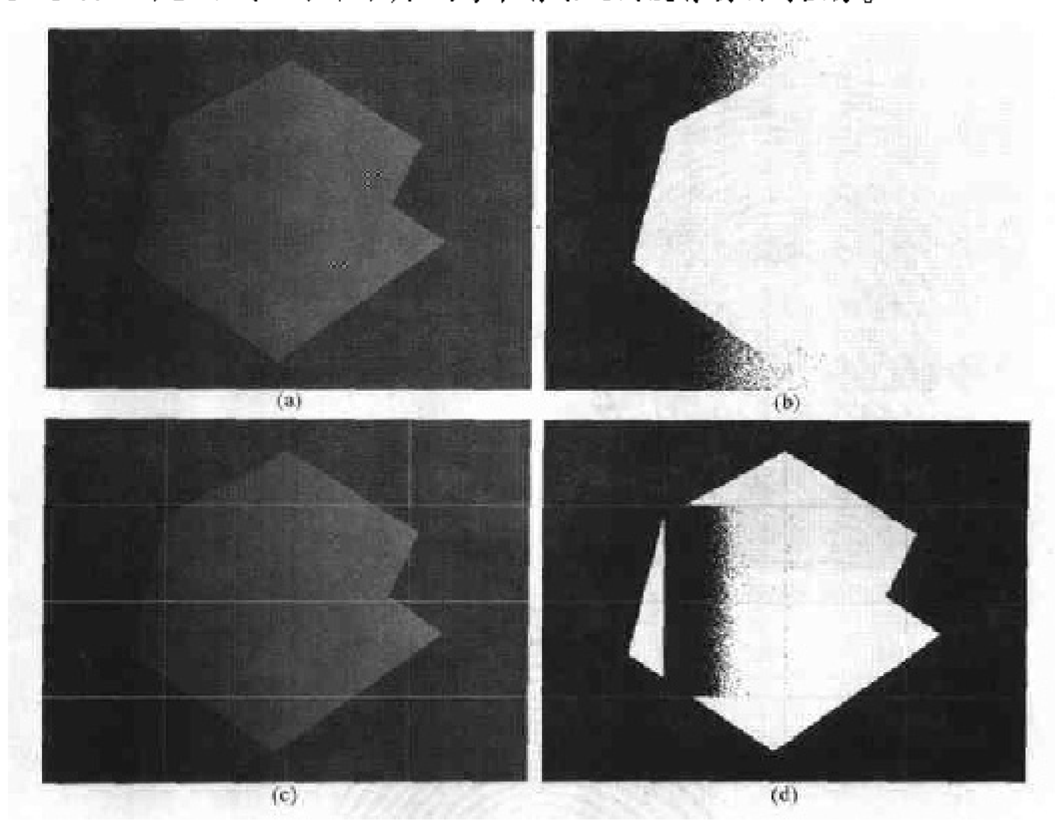


图10.30 (a)原图,(b)经全局门限处理后的结果,(c)被细分为单个子图的图像,(c)经自适应处理后得到的结果

图10.31(d)显示了进一步分割为更小的子图像时失败的子图像。图10.31(e)显示了左上角子图像的直方图。这个子图包含了从对象到背景的过渡区域。这个更小的子图像有一幅清晰的双峰直方图,并且很容易分割。实际上,这种情况就如图10.31(f)所示。这幅图还显示了其他小子图像的分割。所有这些子图像都有近于单峰的直方图,它们的平均灰度级更接近于对象而不是背景,所以它们都被归类为对象。这幅图像给读者留下了一种解决方案,说明通过将图10.30(a)中的整幅图像细分为图10.31(d)中所示大小的子图像可以对图像进行更为精确的分割。

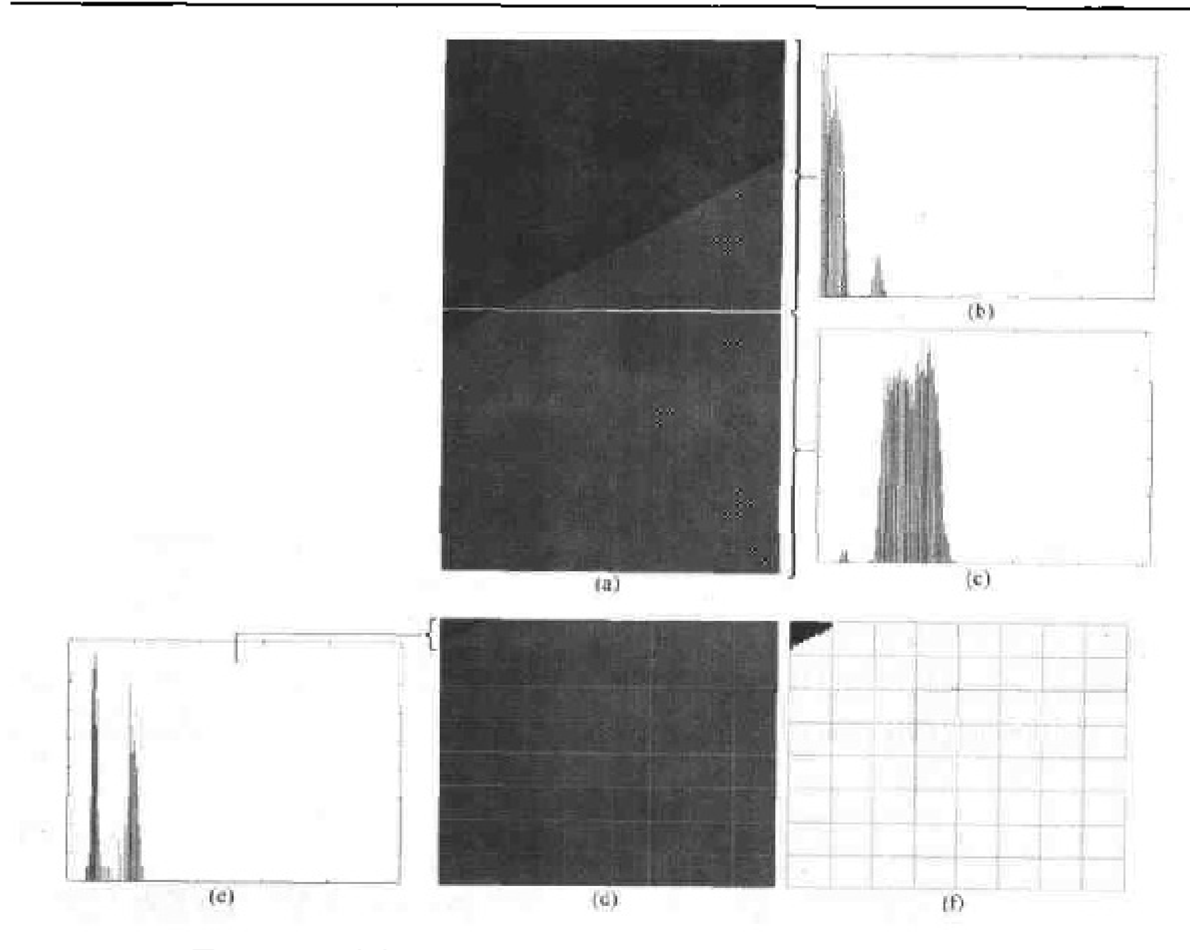


图 10.31 (a)来自图 10.30 的进行了适当和不适当分割的子图,(b)~(c)对应的直方图,(d)对进行了不适当分割的子图像进一步的细分,(e)左上角的小子图的直方图,(f)对(d)进行自适应分割的结果

10.3.5 最佳全局和自适应门限

本节中,我们讨论一种产生最小平均分割误差的估计门限的方法。正如所说明的那样,这种方法被应用于解决实际使用门限处理方法的过程中经常出现的几个重要问题。

假设一幅图像仅包含两个主要的灰度级区域。令 z 表示灰度级值。我们将这些值看做随机量,并且它们的直方图可以被看做它们的概率密度函数(PDF)的估计 $p(z)$。这个总密度函数是两个密度的和或混合。一个是图像中亮区域的密度,另一个是暗区域的密度。再有,混合参数与相关的亮和暗区域是成比例的。如果密度的表达式已知或进行了假设,则它能够确定一个最佳门限(具有最低的误差)将图像分割为两个可区分的区域。

图 10.32 说明了两个概率密度函数。假设两个 PDF 中较大的一个对应于背景的灰度级,而较小的一个描述了图像中对象的灰度级。描述图像中整体灰度级变化的混合概率密度函数是:

$$p(z) = P_1 p_1(z) + P_2 p_2(z) \tag{10.3.5}$$

这里 P_1 和 P_2 是两类像素出现的概率;即 P_1 是随机像素的概率,该像素是具有值 z 的对象像素(一个数)。同样,P_2 是像素属于背景像素的概率。我们假设任何给定的像素不是属于对象就是属于背景,使得:

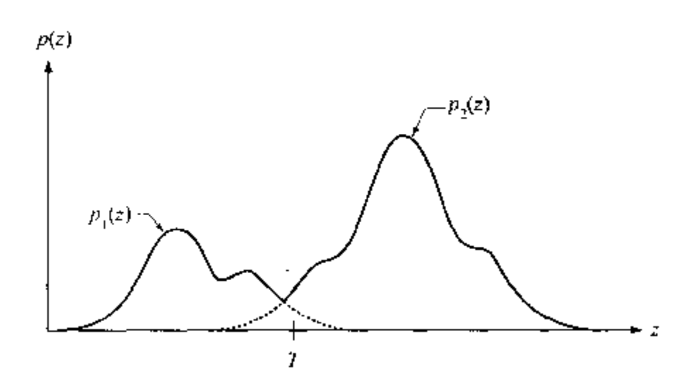


图 10.32 一幅图像中两个区域的灰度级概率密度函数

$$P_1 + P_2 = 1 \tag{10.3.6}$$

一幅图像通过将像素灰度级值大于门限 T(见图 10.32)的像素归类为背景进行分割。所有其他像素归类为对象。主要的目的是选择一个 T 值,使得在决定一个给定的像素是属于对象还是背景时的平均出错率降至最小。

在区间$[a,b]$内取值的随机变量的概率是它的概率密度函数从 a 到 b 的积分,即在这两个上下限之间PDF曲线围住的面积。因此,将一个背景点当做对象点进行分类时,错误发生的概率为:

$$E_1(T) = \int_{-\infty}^{T} p_2(z)\mathrm{d}z \tag{10.3.7}$$

这是在曲线 $p_2(z)$下方位于门限左边区域的面积。同样,将一个对象点当做背景点进行分类时,错误发生的概率为:

$$E_2(T) = \int_{T}^{\infty} p_1(z)\mathrm{d}z \tag{10.3.8}$$

这是在曲线 $p_1(z)$下方位于 T 右边区域的面积。出错率的整体概率是:

$$E(T) = P_2 E_1(T) + P_1 E_2(T) \tag{10.3.9}$$

请注意公式是如何用对象或背景像素出现的概率对量 E_1 和 E_2 进行加权的。同时请注意它们的脚标是相对应的。公式的这种形式很好解释。例如,考虑背景点从不出现的极端情况,此时 $P_2 = 0$。因为背景点从不出现,所以将背景点归类为对象点(出错率为 E_1)的概率在整体出错率(E_1)中的分布可以不加考虑。此时的出错率可以通过用 $P_2 = 0$ 乘以 E_1 表示。如果背景和对象点以等可能性出现,则权值 $P_1 = P_2 = 0.5$。

要找到出错最少的门限值需要将 $E(T)$对 T 求微分(使用莱布尼兹法则)并令微分式等于0。结果是:

$$P_1 p_1(T) = P_2 p_2(T) \tag{10.3.10}$$

这个等式解出 T,即为最佳门限。注意如果 $P_1 = P_2$,则最佳门限位于曲线 $p_1(z)$和 $p_2(z)$的交点处(见图 10.32)。

得到一个 T 的分析表达式需要知道两个PDF的等式。在实践中并不是总可以对这两个密度进行估计。通常做法是利用参数比较易于得到的密度。这种情况下使用的主要密度之一是高斯密度。高斯密度可以用两个参数完全描述:均值和方差。此时,

$$p(z)=\frac{P_1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(z-\mu_1)^2}{2\sigma_1^2}}+\frac{P_2}{\sqrt{2\pi}\sigma_2}e^{-\frac{(z-\mu_2)^2}{2\sigma_2^2}} \tag{10.3.11}$$

这里 μ_1 和 σ_1^2 分别是某一类像素(即对象)的高斯密度的均值和方差,μ_2 和 σ_2^2 分别是另一个类的均值和方差。将这个方程用于解式(10.3.10)的一般解得到下列门限 T 的解:

$$AT^2+BT+C=0 \tag{10.3.12}$$

这里:

$$\begin{aligned}A&=\sigma_1^2-\sigma_2^2\\B&=2(\mu_1\sigma_2^2-\mu_2\sigma_1^2)\\C&=\sigma_1^2\mu_2^2-\sigma_2^2\mu_1^2+2\sigma_1^22\sigma_2^2\ln(\sigma_2P_1/\sigma_1P_2)\end{aligned} \tag{10.3.13}$$

由于二次方程有两个可能的解,所以要得到最佳解需要两个门限值。

如果方差都相等,$\sigma^2=\sigma_1^2=\sigma_2^2$,则单一的门限就足够了:

$$T=\frac{\mu_1+\mu_2}{2}+\frac{\sigma^2}{\mu_1-\mu_2}\ln\left(\frac{P_2}{P_1}\right) \tag{10.3.14}$$

如果 $P_1=P_2$,最佳门限是均值的平均数。如果 $\sigma=0$,结果也一样。对其他已知形式的密度,比如瑞利和log-normal密度,决定最佳门限的方法相似。

与为 $p(z)$假设一种函数形式不同,一种最小均方误差方法能由图像的直方图估计一幅图像复合灰度级的PDF。例如,(连续的)混合密度 $p(z)$和(离散的)图像直方图 $h(z_i)$之间的均方误差为:

$$e_{ms}=\frac{1}{n}\sum_{i=1}^{n}[p(z_i)-h(z_i)]^2 \tag{10.3.15}$$

这里 n 点直方图是假设的。对全部密度进行估计的主要原因是,要确定PDF中是否存在主模式。例如,两个主模式典型地指出在图像(或区域)中存在边缘。这种方式优于计算图像的PDF。

一般来讲,在分析上确定能使均方误差最小的参数不是一件简单的事。甚至对高斯分布,令偏导数为0的简单计算也会导出一组联立超越方程,通常这种方程只能通过诸如共轭梯度或用于解联立非线性方程组的牛顿法一类的数学方案才能解决。

例10.13 使用最佳门限处理对图像进行分割

接下来的是在图像处理中最早将最佳门限应用于分割的例子之一(也是最有启发性的例子之一)。在此处,这个例子非常有意思,因为它说明了如何通过使用预处理技术改善图像分割的效果,这一预处理技术以我们在讨论图像增强时所展开的方法为基础。另外,这个例子也说明了局部直方图估计和自适应门限处理的用处。这个例子中的一般问题是要在心血管造影照片(已经注射了造影剂的X射线图像)中自动地描绘心室边界。此处讨论的方法由Chow和Kaneko[1972]发明,目的是描绘心脏的左心室边界。

在分割之前,所有的图像要进行如下的预处理:(1)每个像素用取对数的方法(见3.2.2节)进行映射以减少由于辐射吸收引起的指数效应;(2)为了去除两幅图像中的脊柱,从注入造影剂后获取的图像中减去使用造影剂之前得到的图像(见3.4.1节);(3)将几张血管造影片相加以便减少随机噪声(见3.4.2节)。图10.33显示了预处理前后的心血管造影照片(在下面的段落中将对 A 和 B 区域进行解释)。

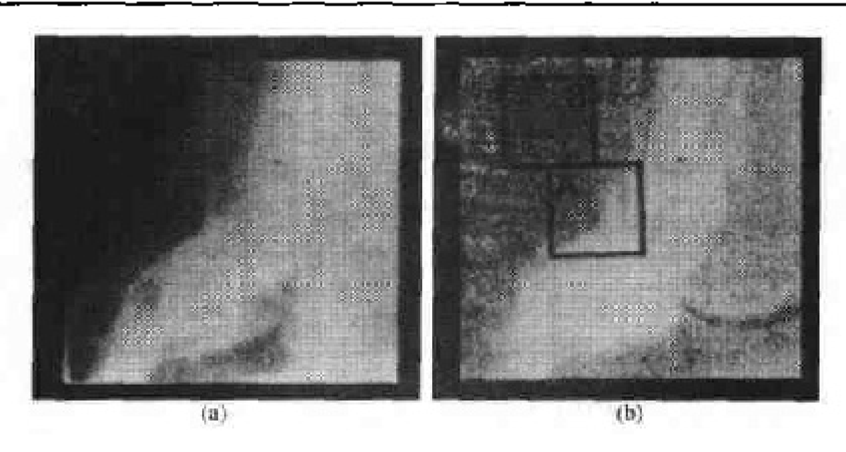


图 10.33 预处理前后的心血管造影照片(见 Chow 和 Kaneko 的著述)

为了计算最佳门限,每幅经过预处理的图像通过一个 7×7 大小且每幅图像有50%重叠率的网格被分成 49 个区域(此处所有的原始图像尺寸均为 256×256 像素)。49 个重叠区域的每一个都包含 64×64 像素。图 10.34(a)和 10.34(b)是图 10.33(b)中标记为 A 和 B 的区域的直方图。注意,区域 A 的直方图很清楚是双峰的,表示其中存在边界。然而,区域 B 的直方图是单峰的,表示没有具有显著区别的区域。

对所有 49 个区域进行计算之后,进行双峰检测以便去除单峰的直方图。剩下的直方图用双峰的高斯密度曲线进行拟合[见式(10.3.11)],使用共轭梯度爬山法将式(10.3.15)给出的误差函数降至最小。图 10.34(a)中的 × 和 ○ 是直方图中以黑色点表示的适配符号。最佳门限通过使用式(10.3.12)和式(10.3.13)得到。

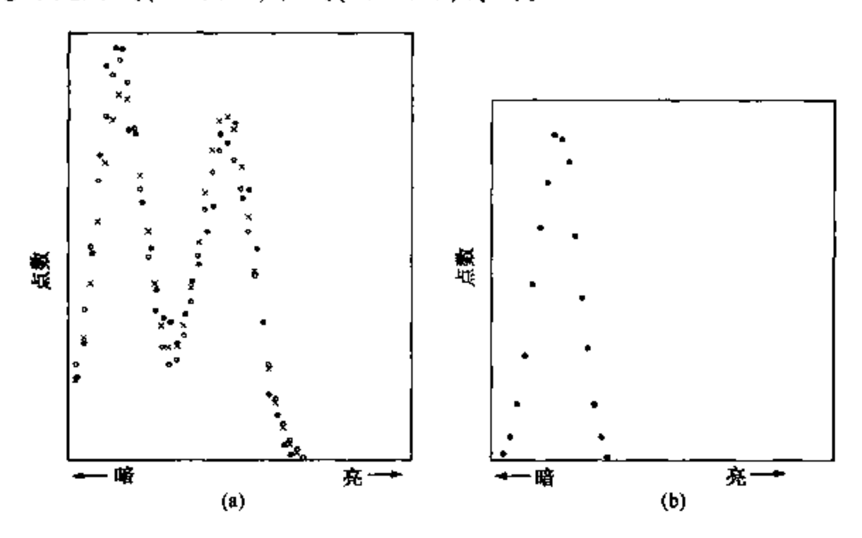


图 10.34 图 10.33(b)中的(a)区域 A 的直方图,(b)区域 B 的直方图(见 Chow 和 Kaneko 的著述)

在处理过程的这个阶段,只有具有双峰直方图的区域设置了门限。可以通过内插的方法对余下区域设置门限。然后,使用邻近的门限值对每个点逐个运用二级插值法,以便在整个过程结束时图像中的每个点都有一个门限。最后,使用下列规则对每个像素进行二元判定:

$$f(x,y)=\begin{cases}1 & f(x,y)\geqslant T_{xy}\\ 0 & \text{其他}\end{cases}$$

这里 T_{xy} 是赋予图像中的位置(x,y)的门限[注意这些都是自适应门限。因为它们的值取决于空间坐标]。用二值图像的梯度法可以得到边界。图10.35显示了叠加在原图上的边界。考虑到图像的可变性和复杂性,这个过程得到了极好的分割结果。

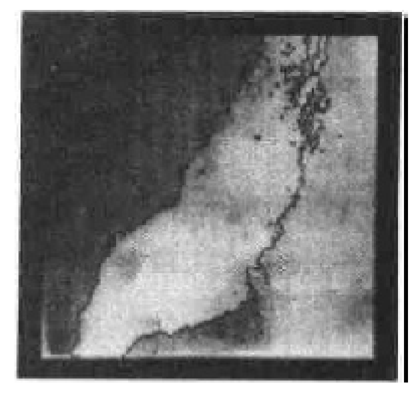


图10.35 叠加了边界的心血管造影图(见Chow和Kaneko的著述)

10.3.6 利用边界特性改进直方图和局部门限处理

根据前五节的讨论,可以很明显地看到:如果直方图的尖峰很高、很窄、具有对称性且被很深的波谷割开,则找到"好"的门限的机会会有相当大的提高。一种改进直方图形状的方法是只考虑那些位于对象和背景之间边缘上的或在边缘附近的像素。使用这种方法可以马上明显地看到,直方图对于对象和背景大小的依赖性变小了。例如,一幅由很小的对象和很大的背景区域组成的图像(或正好相反)的直方图中,一个大的尖峰是图像的主要部分,因为某一类像素数目高度集中。图10.30和图10.31很好地说明了这种条件对分割性能的影响。

如果只使用位于对象和背景间边缘上的或附近的像素,得到的直方图会有同样高度的尖峰。另外,给定像素位于对象内的概率会与像素位于背景内的概率大体相等,由此改进了直方图尖峰的对称性。最后,正如将在下面段落中说明的那样,使用满足某些简单度量的像素会趋向于加深直方图尖峰之间的波谷,这些简单度量是以梯度和拉普拉斯算子为基础的。

刚才谈到的这种方法存在的主要问题是,隐含了认为对象和背景之间的边缘为已知的这一假设。很明显,当在对象和背景之间找到的分割线正好全部分割的时候,这一信息在分割中是没有用的。然而,根据10.1.3节中的讨论,一个像素是否处在边缘上的迹象可通过计算它的梯度得到。另外,使用拉普拉斯算子还可以知道给定的像素是在边缘亮的一边还是暗的一边。拉普拉斯算子的平均值在边缘的过渡部分处为0(见图10.6),所以,实际上,根据梯度或拉普拉斯算子准则选择的像素构成的直方图波谷,可以被认为是稀疏分布的。这一特性产生前边讨论过的十分希望的深谷。

在图像中任意点(x,y)的梯度∇f由式(10.1.4)或式(10.1.12)给出。同样,拉普拉斯算子$\nabla^2 f$由式(10.1.14)或式(10.1.15)给出。这两个量可以用于生成一个三级图像,如下所示:

$$s(x,y)=\begin{cases}0 & \nabla f < T \\ + & \nabla f \geq T, \quad \nabla^2 f \geq 0 \\ - & \nabla f \geq T, \quad \nabla^2 f < 0\end{cases} \tag{10.3.16}$$

这里符号0,+和-表示任意3个不同的灰度级,T是门限,并且计算每个点(x,y)的梯度和拉普拉斯值。对于在亮背景中的暗对象,参考图10.6,使用式(10.3.16)生成图像$s(x,y)$:其中(1)所有不在边缘上的像素(由∇f小于门限T确定)标记为0;(2)所有在边缘"暗"的一边的像素标记为+;(3)所有在边缘"亮"的一边的像素标记为-。式(10.3.16)中的符号+和-对于处在"暗"背景中的"亮"对象是相反的。图10.36说明了由式(10.3.16)产生的亮背景下所写的下划线的暗图像打印标记。

从这一过程中得到的信息可以用于生成经过分割的二值图像,图中的1对应于有意义的

对象,0对应于背景。从亮背景到暗对象的过渡(沿着水平或垂直扫描线)必须以 $s(x,y)$ 中一个“-”号后跟一个“+”号的形式为特点。对象内部的像素不是标记为“0”就是标记为“+”。最后,从对象反向向背景的过渡用一个“+”号后跟一个“-”号的形式表示。因此,包含对象部分的一条水平或垂直扫描线具有如下结构:

$$(\cdots)(-,+)(0\text{或}+)(+,-)(\cdots)$$

这里(…)表示-,+和0的任意组合。最内层的括号包含着对象点,并被标记为1。沿着相同的扫描线除了那些被(-,+)和(+,-)包围着的(0或+)的序列以外,所有其他点都标记为0。

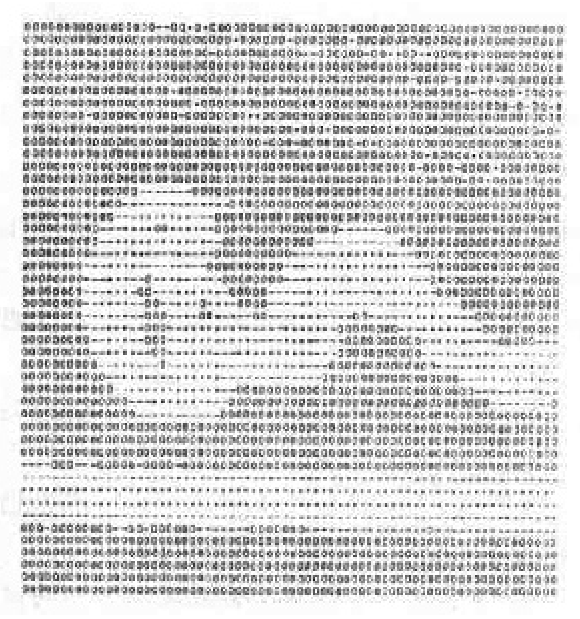


图10.36 使用式(10.3.16)得到的用打字机打出的手写字母编码的图像(由IBM提供)

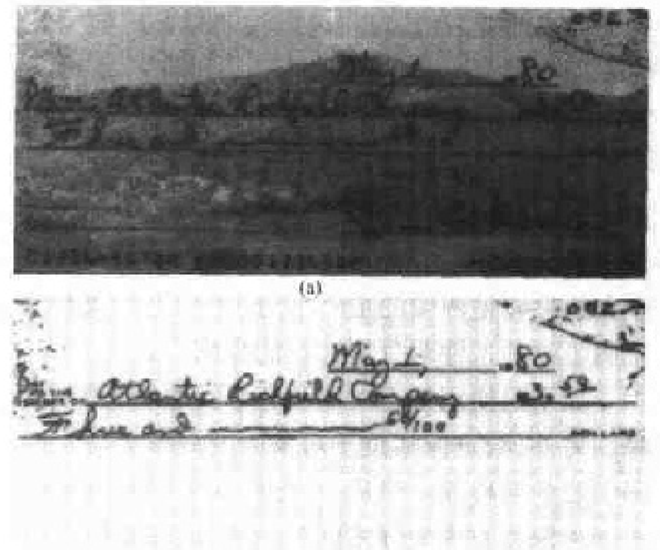


图10.37 (a)原图像,(b)使用局部门限分割过的图像(由IBM提供)

例10.14 用局部门限的图像分割

图10.37(a)显示了一张普通的银行支票。图10.38显示了以梯度值为函数并且梯度值大于5的像素的直方图。注意,此直方图有两个对称的、近似等高的主模式,这两个主模式由一个区分波谷割开。最后,图10.37(b)显示了使用式(10.3.16)和位于波谷中点附近的 T 值的条件下得到的分割图像。结果用刚才提到的分析步骤产生二值图像。注意,这个例子说明的是式(10.3.1)定义的局部门限处理,因为 T 的值取决于具有局部特性的梯度和拉普拉斯算子的直方图。

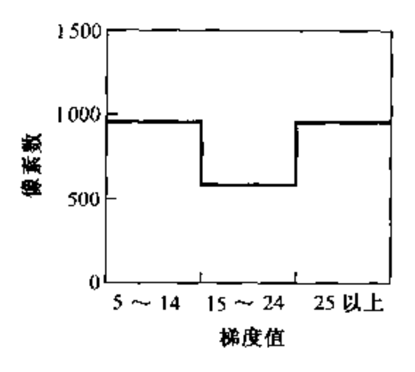


图10.38 梯度值大于5的像素的直方图(由IBM提供)

10.3.7 基于不同变量的门限

迄今为止我们关注的是门限处理的灰度级。在某些情况下,传感器可以产生不止一个在图像中描述每一个像素的可利用的变量,因此,允许进行多谱段门限处理。如6.7节中较详细

讨论的，彩色成像是一个很好的例子，它的每个像素分别由3个RGB值描述。此时，可以构造一个三维“直方图”。基本过程类似于一个变量时使用的方法。例如，对一幅有3个变量的图像（RGB分量），每个像素有16种可能的灰度级，这就构成了一个16×16×16的网格（立方体）。填入每个立方体单元内的数字是RGB分量的值与定义特殊单元位置的坐标相一致的像素编号。每个记入项被图像中像素的总数除，形成一个归一化的直方图。

现在，门限处理的概念变成在三维空间内寻找点的聚簇的过程。例如，假设在直方图中找到有效点簇 K，我们可以对RGB分量值接近某一个簇的像素赋予一个任意值（如白色的值）而对图像中的其他像素赋予其他的值（如，黑色的），通过这种方式将图像进行分割。这一概念很容易推广至具有更多的分量和更多簇的情况。主要困难在于随着变量数目的增加，簇的搜索也变成一种更复杂的工作。可以建立簇的搜索方法。比如，Duda，Hart和Stork[2001]以及Tou和Gonzalez[1974]的著述中都有介绍。

例10.15　多谱门限处理

图10.39(a)中所示的图像是一幅彩色照片的单色图像。原彩色图是由16位RGB图像组成的。围巾是鲜红色，头发和脸部的颜色很浅且在光谱特性上与窗户和其他的背景特征迥然不同。

图10.39(b)是用直方图中对应于脸部色调的一个簇进行门限处理后得到的。注意，在单色图中窗户的灰度级接近于头发的灰度级值。

由于多频谱特性用于分割这两个不同区域，因而在此图中窗户没有显示出来。图10.39(c)是对接近于红色轴的簇进行门限处理得到的图像。此时，只有围巾和花（也是红色的）的一部分出现在分割结果中。用于得到这两幅图像的门限是某个单元的距离。因此，在该问题中对任何像素，只要它的分量在某个单元之外并接近簇的中心就把它归类为背景（黑色）。各分量在这个单元内的像素被编码为白色。



图10.39　(a)以单色图显示的彩色照片，(b)用接近于脸部的色调对像素进行分割，(c)红色分量的分割结果

如6.7节中提到的，色彩分割是以第6章中介绍的所有颜色模型为基础的。例如，色调和饱和度是很多自动监测应用中很重要的性质。这些性质在试图模拟人所具有的功能时特别的重要。比如检测水果是否成熟或检测制造的货物。如第6章中所描述的，色调、饱和度、强度模型（Hue，Saturation，Intensity，HSI）对这类应用是很理想的，因为这种模型非常接近于人类对颜色的感觉方式。使用颜色信号的色调和饱和度分量进行分割的方法具有特别的吸引力，因为它用到的二维数据簇比RGB分割需要的三维簇要易于分析。

10.4 基于区域的分割

分割的目的是将图像划分为不同区域。在 10.1 节和 10.2 节中,根据区域间灰度不连续通过搜寻区域之间的边界来处理这一问题,而 10.3 节中分割是通过用以像素性质的分布为基础的门限来进行的,比如灰度级的值或颜色。在本节中,我们讨论以直接找寻区域为基础的分割技术。

10.4.1 基本公式

令 R 表示整幅图像区域。可以将分割看做将 R 划分为 n 个子区域 $R_1, R_2, \cdots, R_n$ 的过程:

(a) $\bigcup_{i=1}^{n} R_i = R$。

(b) R_i 是一个连通的区域,$i = 1,2,\cdots,n$。

(c) $R_i \cap R_j = \varnothing$,对所有的 i 和 j,$i \neq j$。

(d) $P(R_i) = \text{TRUE}$,对于 $i = 1,2,\cdots,n$。

(e) $P(R_i \cup R_j) = \text{FALSE}$,对于 $i \neq j$。

这里,$P(R_i)$是定义在集合 R_i 的点上的逻辑谓词,$\varnothing$是空集。

条件(a)说明分割必须是完全的;即,每个像素必须属于一个区域。条件(b)要求区域中的点必须与某个预定义的准则相联系(见 2.5.2 节关于连通性的内容)。条件(c)说明不同区域必须是不相交的。条件(d)涉及在分割区域内的像素必须满足的性质——例如如果所有 R_i 内的像素有相同的灰度级,则 $P(R_i) = \text{TRUE}$。最后,条件(e)说明区域 R_i 和 R_j 对于谓词 P 是不同的。

10.4.2 区域生长

正如它的名字所暗示的,区域生长是一种根据事前定义的准则将像素或子区域聚合成更大区域的过程。基本的方法是以一组"种子"点开始将与种子性质相似(诸如灰度级或颜色的特定范围)的相邻像素附加到生长区域的每个种子上。

如在例 10.16 中显示的那样,通常根据所解决问题的性质而选择一个或多个起点。当一个先验信息无效时,这一过程将对每个像素计算相同的特性集,最终,这个特性集在生长过程中用于将像素归属某个区域。如果这些计算的结果呈现了不同簇的值,则那些由于自身的性质而处在这些簇中心附近的像素可以作为种子。

相似性准则的选择不仅取决于面对的问题,还取决于有效图像数据的类型。例如,对地观测卫星成像非常依赖颜色的使用。如果没有彩色图像本身固有的可用信息,这个问题会变得非常棘手,甚至无法解决。如果图像是单色的,必须用一组基于灰度级和空间性质的描绘子(如矩或纹理)对区域进行分析。我们将在第 11 章中讨论对区域特性有用的描绘子。

如果有关连通性和相邻性的信息没有用于区域生长过程,则单个的描绘子会产生错误的结果。例如,仅用 3 个不同的灰度级值将任意一个范围内的像素可视化。具有相同灰度级的

像素组成一个"区域",而不用考虑它们的连通性。此时会生成一个对于我们现在讨论的内容毫无意义的分割结果。

区域生长的另一个问题是用公式描述一个终止规则。基本上,在没有像素满足加入某个区域的条件时,区域生长就会停止。像灰度级、纹理和颜色准则都是局部性质,都没有考虑到区域生长的"历史"。其他增强了区域生长算法处理能力的准则利用了待选像素和已加入生长区的像素间的大小、相似性等概念(比如待选像素的灰度级和生长区域的平均灰度级之间的比较),以及生长区域的形状。这些类型的描绘子的使用是以假设能得到预期结果的模型至少有一部分有效为基础的。

例 10.16　区域生长在焊缝检测中的应用

图 10.40(a)显示了一幅焊缝(水平深色区域)的 X 射线图像,图像中含有几条裂缝和孔隙(横向通过图像中间部位的亮白色条纹)。我们希望用区域生长的方法将有缺陷的焊接区域分离出来。这些分割的特性可以应用于检测、历史研究的数据库,以及控制自动焊接系统及其他很多方面。

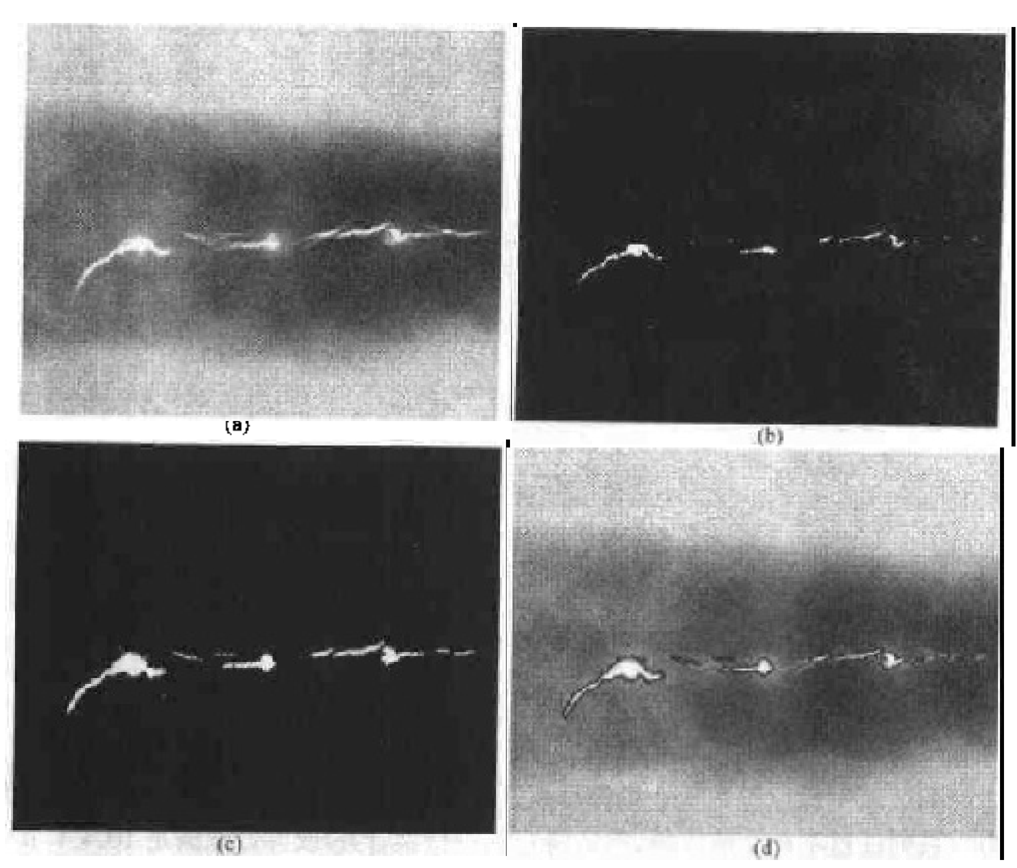


图 10.40　(a)显示有缺陷的焊缝的图像,(b)种子点,(c)区域生长的结果,(d)对有缺陷的焊缝区域进行分割得到的边界(用黑色表示)(原图由X-TEK系统有限公司提供)

工作第一步是决定初始的种子点。在这个应用中,我们知道表现有缺陷的焊缝的像素趋向于取允许的最大数字值(此时为 255)。根据这一信息,我们选择所有值为 255 的像素作为起点。图 10.40(b)中显示了从原图中提取出来的点。注意,很多点聚集为种子区。

下一步是为区域生长选定准则。在这个特定的例子中,我们为一个像素是否能添加到某

一区域制定两个标准:(1)任何像素和种子之间的灰度级绝对差必须小于65,这个数字是根据图10.41中显示的直方图得来的,它表示了255和从左边开始第一个主要的波谷的位置所代表的值之间的差,这个波谷表示在暗的焊缝区域中具有的最高的灰度级;(2)要添加入某一个区域的像素必须与此区域中至少一个像素是8连通的。如果某个像素被发现与多于一个区域相联系,就将这些区域合并在一起。

图10.40(c)显示了从图10.40(b)中的种子开始并使用上一段中定义的准则得到的区域。将这些区域的边界叠加到原图中[图10.40(d)]显示出:区域生长过程确实以可以接受的精确度对有缺陷的焊缝进行了分割。值得注意的有趣的地方是这一过程无须指定任何终止规则。因为,区域生长的准则足以分离出我们关心的特征。

在10.3.1节中及图10.26(b)中讲过具有多峰分布直方图的问题最好使用基于区域的方法解决。图10.41中显示的直方图是具有"清晰的"多峰直方图的很好的例子。这幅直方图和例10.16的结果证明了这样的结论:即使具有更规范的直方图,多级门限处理也是一个困难的命题。根据例10.16的结果我们很直观地就可以看出这类问题使用仅以灰度级为基础的自动门限选择方法是无法得到有效解决的。应用连通性是解决这类问题的基础。

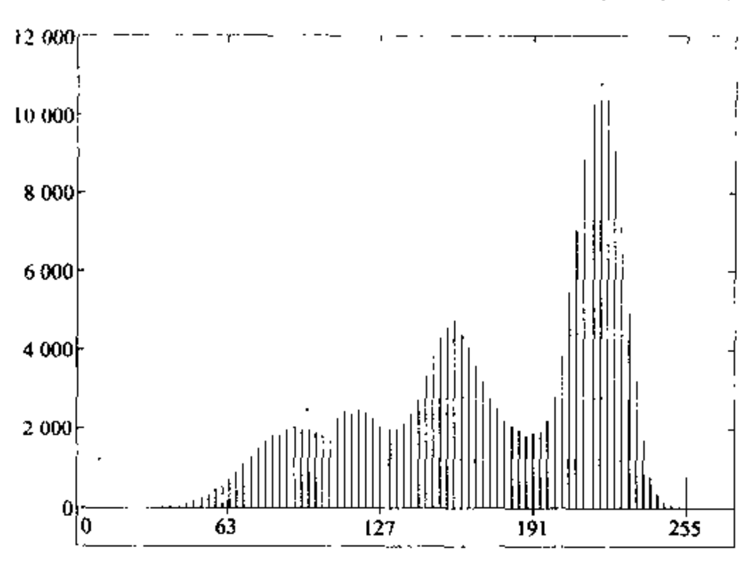


图10.41 图10.40(a)的直方图

10.4.3 区域分离与合并

刚才讨论的区域生长过程是从一组种子点开始的。另一种可作为替换的方法是在开始时将图像分割为一系列任意不相交的区域,然后将它们进行聚合并/或拆分以满足10.4.1节中阐述的条件。我们将在下面对这种通过反复进行拆分和聚合以满足限制条件的算法展开讨论。

令 R 表示整幅图像区域并选择一个谓词 P。对 R 进行分割的一种方法是反复将分割得到的结果图像再次分为四个区域,直到对任何区域 R_i,有 $P(R_i)=\text{TRUE}$。这里是从整幅图像开始。如果 $P(R)=\text{FALSE}$,就将图像分割为4个区域。对任何区域如果 P 的值是FALSE,就将这4个区域的每个区域再次分别分为4个区域,如此不断继续下去。这种特殊的分割技术用所谓的四叉树形式表示最为方便(就是说,每个非叶子节点正好有4个子树),这正如图10.42中说明的树那样。注意,树的根对应于整幅图像,每个节点对应于划分的子部分。此

时，只有 R_4 进行了进一步的再细分。

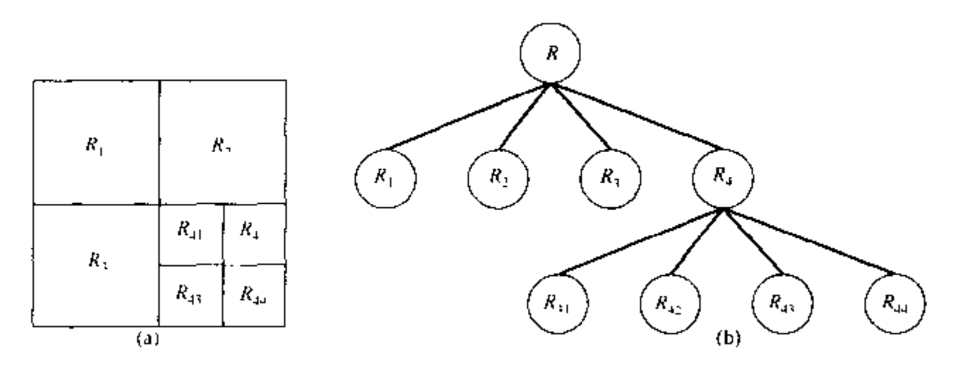


图 10.42 (a)被分割的图像，(b)对应的四叉树

如果只使用拆分，最后的分区可能会包含具有相同性质的相邻区域。这种缺陷可以通过进行拆分的同时也允许进行区域聚合来得到矫正。满足 10.4.1 节中提出的限制条件仅需要将相邻区域进行聚合，但聚合的不同区域中相互连接的像素必须满足谓词 P。就是说，只有在 $P(R_j \cup R_k)=\text{TRUE}$时，两个相邻的区域 R_j 和 R_k 才能聚合。

前面的讨论可以总结为如下过程。在反复操作的每一步，我们需要做：

1. 对于任何区域 R_i，如果 $P(R_i)=\text{FALSE}$，就将每个区域都拆分为 4 个相连的象限区域。
2. 将 $P(R_j \cup R_k)=\text{TRUE}$ 的任意两个相邻区域 R_j 和 R_k 进行聚合。
3. 当再无法进行聚合或拆分时操作停止。

可以对前面讲述的基本思想进行几种变化。例如，一种可能的变化是开始时将图像拆分为一组图像块。然后对每个块进一步进行上述拆分，但聚合操作开始时受只能将 4 个块并为一组的限制。这 4 个块是四叉树表示法中节点的后代且都满足谓词 P。当不能再进行此类聚合时，这个过程终止于满足步骤 2 的最后的区域聚合。在这种情况下，聚合的区域可能会大小不同。这种方法的主要优点是对于拆分和聚合都使用同样的四叉树，直到聚合的最后一步。

例 10.17 拆分和聚合

图 10.43(a)显示了一幅简单的图像。如果在区域 R_i 内至少有 80% 的像素具有 $|z_j - m_i| \leq 2\sigma_i$ 的性质，就定义 $P(R_i)=\text{TRUE}$，这里 z_j 是 R_i 内第 j 个像素的灰度级，m_i 是区域 R_i 的灰度级均值，σ_i 是区域 R_i 内的灰度级的标准差。如果在此条件下，$P(R_i)=\text{TRUE}$，则设置 R_i 内的所有像素的值等于 m_i。拆分和聚合使用前述算法的要点完成。将这种技术应用于图 10.43(a)所得结果示于图 10.43(b)。请注意，图像分割效果相当好。示于图 10.43(c)中的图像是通过对图 10.43(a)进行门限处理得到的，门限值选在直方图中两个主要的尖峰之间的中点。经过门限处理，图像中生成的阴影(和叶子的茎)被错误地消除了。

如前面的例子中所使用的属性那样，我们试图使用基于区域中像素的均值和标准差的某些特性对区域的纹理进行量化(见 11.3.3 节中关于纹理的讨论)。纹理分割的概念是以在谓词 $P(R_i)$中使用有关纹理的量度为基础的。就是说，通过指定基于纹理内容的谓词，我们可以使用本节中讨论的任何方法进行纹理分割。

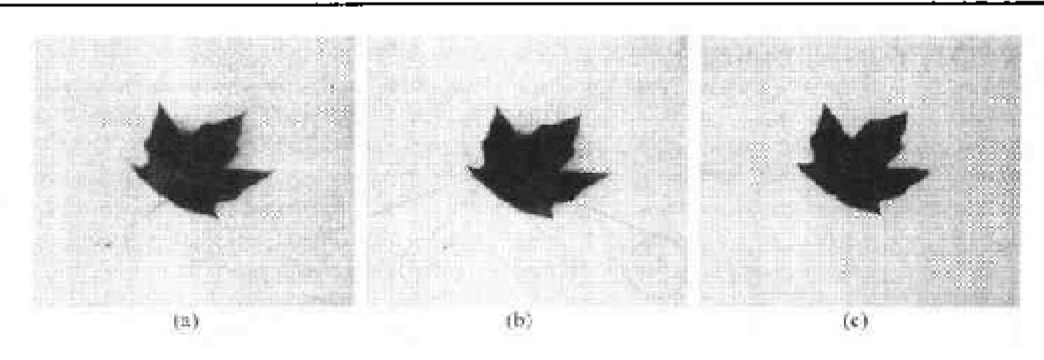


图 10.43　(a)原图,(b)进行拆分和聚合后的结果,(c)对(a)进行门限处理后的结果

10.5　基于形态学分水岭的分割

到现在为止,已经讨论了基于 3 种主要概念的分割方法:(a)间断的检测;(b)门限处理和(c)区域处理。每种方法各有优点(例如,全局门限处理的速度优势)和缺点(如,以灰度级的间断检测为基础的方法需要诸如边线连接等后处理)。本节中,讨论基于所谓的形态学分水岭概念的方法。接下来将要说明,形态学分水岭分割将其他 3 种方法中的许多概念进行了具体化,包括连续的边界分割在内,它生成的分割结果通常更为稳定。这条途径也为在分割过程中结合基于知识的约束提供了一个简单的框架。

10.5.1　基本概念

分水岭概念是以对图像进行三维可视化处理为基础的:其中两个是坐标,另一个是灰度级。对于这样一种“地形学”的解释,我们考虑三类点:(a)属于局部性最小值的点;(b)当一滴水放在某点的位置上的时候,水一定会下落到一个单一的最小值点;(c)当水处在某个点的位置上时,水会等概率地流向不止一个这样的最小值点。对一个特定的区域最小值,满足条件(b)的点的集合称为这个最小值的“汇水盆地”或“分水岭”。满足条件(c)的点的集合组成地形表面的峰线,术语称做“分割线”或“分水线”。

基于这些概念的分割算法的主要目标是找出分水线。基本思想很简单:假设在每个区域最小值的位置上打一个洞并且让水以均匀的上升速率从洞中涌出,从低到高淹没整个地形。当处在不同的汇聚盆地中的水将要聚合在一起时,修建的大坝将阻止聚合。水将只能到达大坝的顶部处于水线之上的程度。这些大坝的边界对应于分水岭的分割线。所以,它们是由分水岭算法提取出来的(连续的)边界线。

这些思想可以用图 10.44 作为辅助做进一步的解释。图 10.44(a)显示了一个简单的灰度级图像。图 10.44(b)是地形图。其中“山峰”的高度与输入图像的灰度级值成比例。为了易于解释,这个结构的后方被遮蔽起来。这是为了不与灰度级值相混淆;三维表达对一般地形学是很重要的。为了阻止上升的水从这些结构的边缘溢出,我们想像将整幅地形图的周围用比最高山峰还高的大坝包围起来。最高山峰的值是由输入图像灰度级可能具有的最大值决定的。

假设在每个区域最小值中打一个洞[如图 10.44(b)中的深色区域],并且让水以均匀的上升速率从洞中涌出,从低到高淹没整个地形。图 10.44(c)说明被水淹没的第一个阶段,这里水用浅灰色表示,覆盖了对应于图中深色背景的区域。在图 10.44(d)和(e)中,我们看到水分别

在第一和第二汇水盆地中上升。由于水持续上升，最终水将从一个汇水盆地中溢出到另一个之中。图 10.44(f)中显示了溢出的第一个征兆。这里，水确实从左边的盆地溢出到右边的盆地，并且两者之间有一个短“坝”(由单像素构成)阻止这一水位的水聚合在一起(在接下来的章节中将讨论坝的构筑)。由于水位不断上升，实际的效果要超出我们所说的。如图 10.44(g)所显示的那样。这幅图中在两个汇水盆地之间显示了一条更长的坝，另一条水坝在右上角。这条水坝阻止了盆地中的水和对应于背景的水的聚合。这个过程不断延续直到到达水位的最大值(对应于图像中灰度级的最大值)。水坝最后剩下的部分对应于分水线，这条线就是要得到的分割结果。

对于这个例子，在图 10.44(h)中显示为叠加到原图上的一个像素宽的深色路径。注意一条重要的性质就是分水线组成一条连通的路径，由此给出了区域之间的连续的边界。

分水岭分割法的主要应用是从背景中提取近乎一致(类似水滴的)的对象。那些在灰度级上变化较小的区域的梯度值也较小。因此，实际上，我们经常可以见到分水岭分割方法与图像的梯度有更大的关系，而不是图像本身。有了这样的表示方法，汇水盆地的局部最小值就可以与对应于所关注的对象的小的梯度值联系起来了。

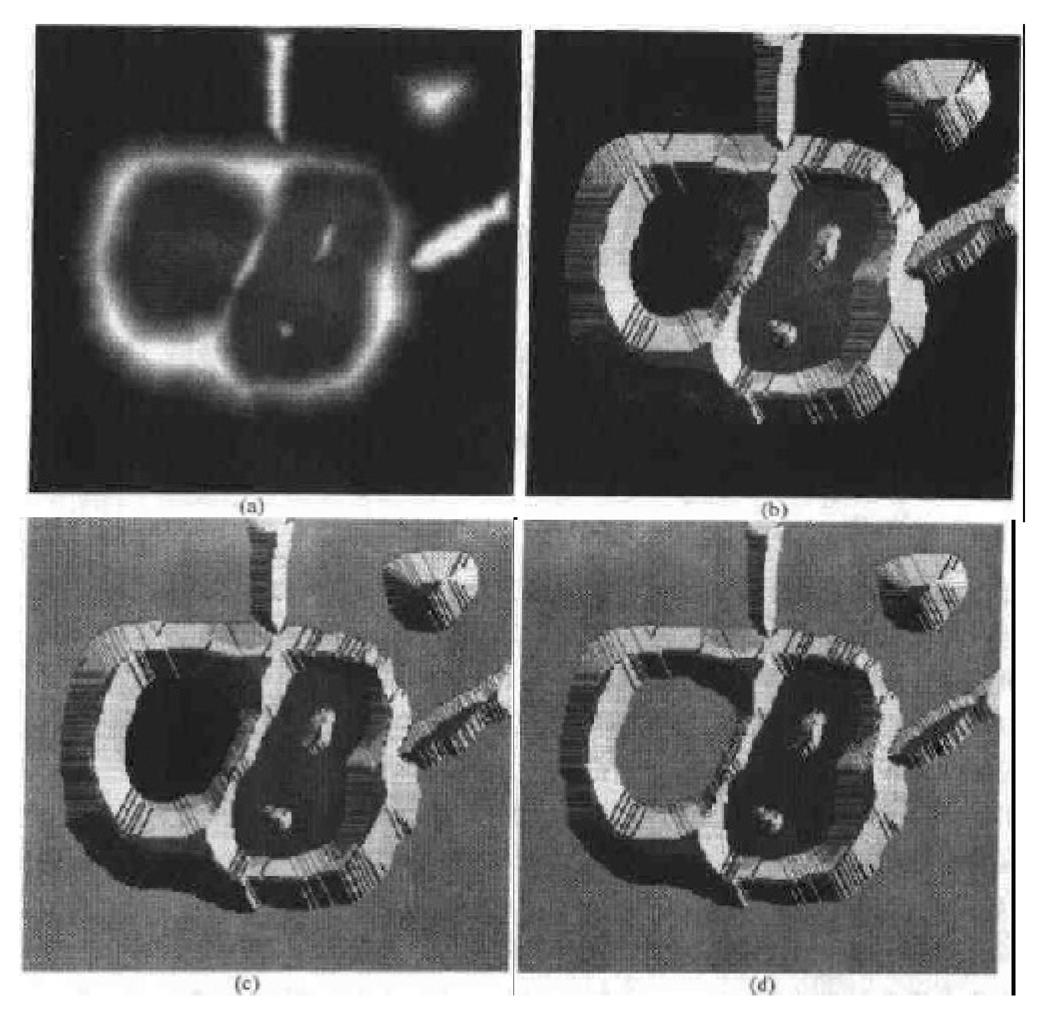


图 10.44　(a)原图像，(b)地形俯视图，(c)～(d)被水淹没的两个不同阶段

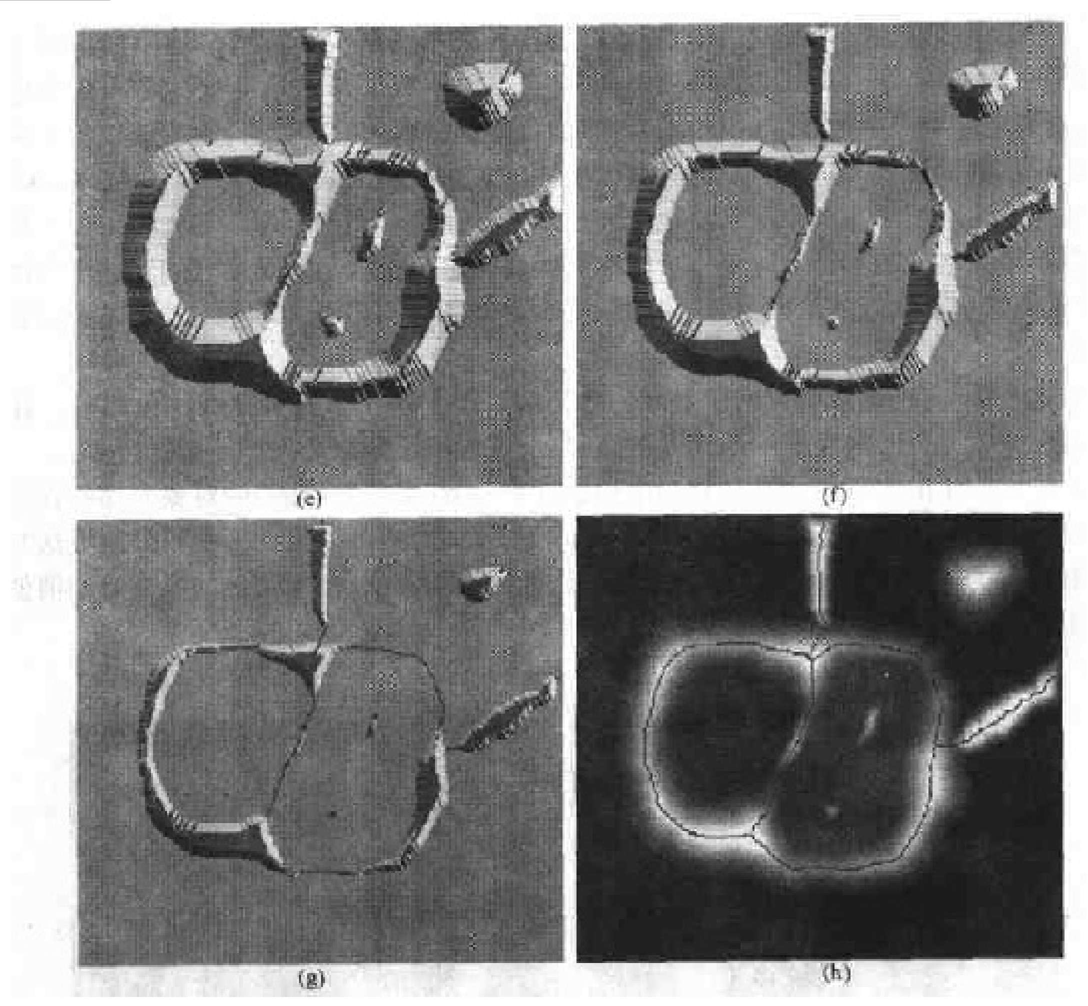


图 10.44(续)　(e)进一步淹没的结果,(f)汇水盆地的水开始聚合(它们之间有一条短水坝),(g)长一些的水坝,(h)最后的分水线(分割)(由 CMM/Ecole des Mines de Paris 的 S.Beucher 博士提供)

10.5.2　水坝构造

在进行讨论之前,让我们考虑一下如何构造分水岭分割方法所需的水坝或分水线。水坝的构造是以二值图像为基础的,这种图像属于二维整数空间 Z^2(见 2.4.2 节)。构造水坝分离二元点集的最简单的方法是使用形态膨胀(见 9.2.1 节)。

图 10.45 说明了如何使用形态膨胀构造水坝的基本点。图 10.45(a)显示了两个汇水盆地的部分区域在淹没步骤的第 $n-1$ 步时的图像。图 10.45(b)显示了淹没的下一步(第 n 步)的结果。水已经从一个盆地溢出到另一个盆地,所以,必须建造水坝阻止这种情况的发生。为了与紧接着要介绍的符号相一致,令 M_1 和 M_2 表示在两个区域极小值中包含的点的坐标集合。然后,将处于汇水盆地中的点的坐标集合与这两个在溢出的第 $n-1$ 个阶段的最小值联系起来,并用 $C_{n-1}(M_1)$和 $C_{n-1}(M_2)$表示。这就是图 10.45(a)中的两个黑色区域。

令这两个集合的联合用 $C[n-1]$表示。图 10.45(a)中有两个连通分量(见 2.5.2 节关于连通分量的部分),而图 10.45(b)中只有一个连通分量。这个连通分量包含着前面的两个分量,用虚线表示。两个连通分量变成一个连通分量的事实说明两个汇水盆地中的水在淹没的

第 n 步聚合了。用 q 表示此时的连通分量。注意，第 $n-1$ 步中的两个连通分量可以通过使用"与"操作($q \cap C[n-1]$)从 q 中提取出来。我们也注意到，属于独立的汇水盆地的所有点构成了一个单一的连通分量。

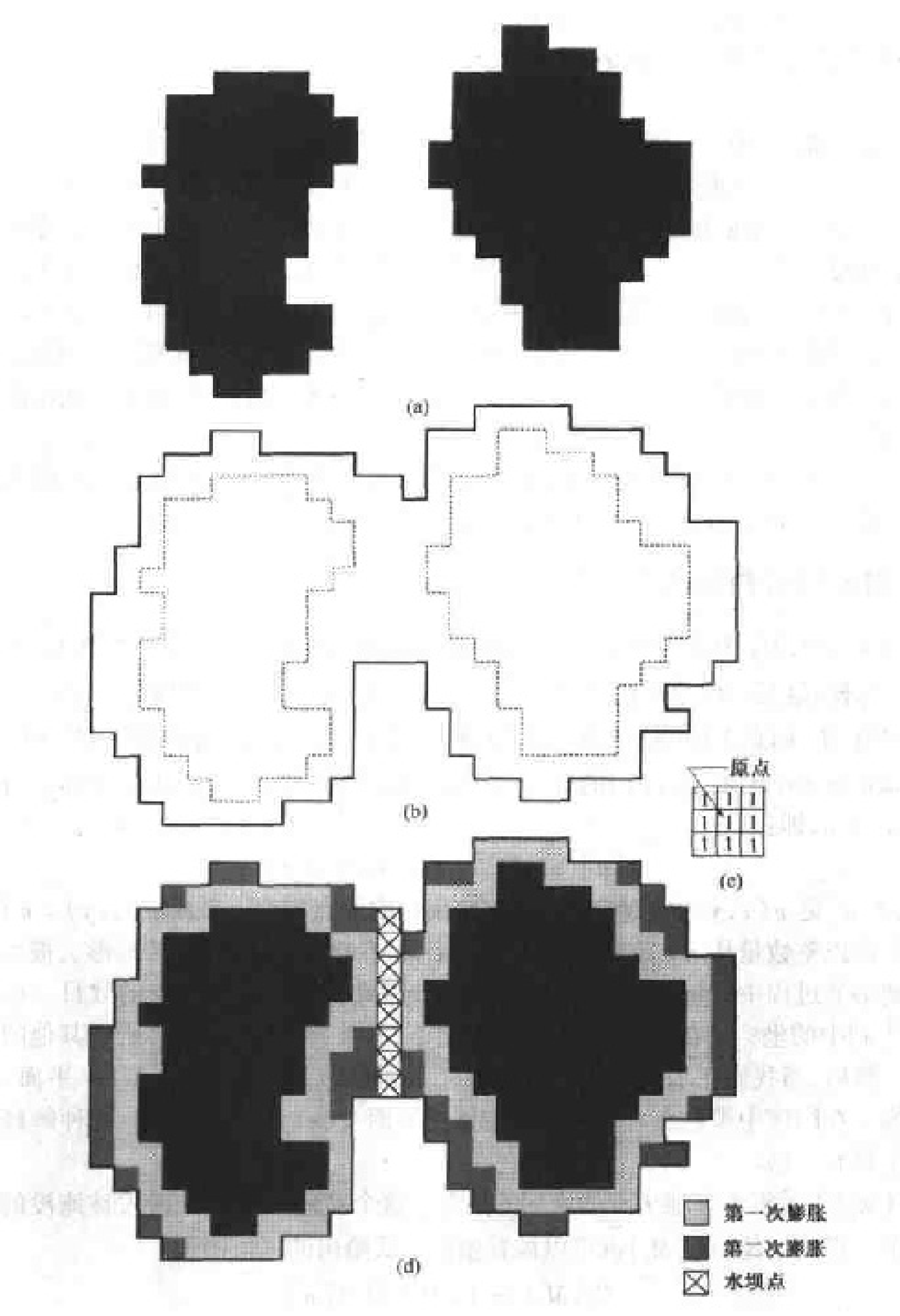


图 10.45 (a)在淹没的第 $n-1$ 个阶段淹没的汇水盆地的两个部分，(b)淹没的第 n 阶段，显示出两个盆地间的水已经溢出(为了显示清晰，水用白色表示而不是黑色)，(c)用于膨胀的结构元素，(d)扩展的结果和水坝的构造

假设图 10.45(a)中的每个连通分量通过使用图 10.45(c)中显示的结构元膨胀，在两个条件下：(1)膨胀受到 q 的约束(这意味着在膨胀的过程中结构化元素的中心只能定位于 q 中)；并且(2)在引起集合聚合的那些点上不能执行膨胀(成为单一的连通分量)。图 10.45(d)显示首轮膨胀(浅灰色表示)使用了每个初始连通分量的边界。注意，在膨胀过程中每个点都满足条件(1)。条件(2)在膨胀处理中没有应用于任何的点；因此，每个区域的边界都进行了均匀的扩展。

在第二轮膨胀中(中等灰度表示)，几个不满足条件(1)的点符合条件(2)时，得到图中显示的断开周界。很明显，只有满足上述两个条件的属于 q 中的点描绘了图 10.45(d)中交叉阴影线表示的一个像素宽度的连通路径。这条路径组成在淹没的第 n 个阶段我们希望得到的水坝。在这个淹没水平上，水坝的构造是由置所有刚好在这条路径上的点的值为比图像中灰度级的最大值还大的值完成的。所有水坝的高度通常设定为 1 加上图像中灰度级最大允许值。这样设定可以阻止在水位不断升高的情况下水越过部分水坝。应该特别注意到的是通过这一过程建立的水坝是连通分量，就是我们希望得到的分割边界。就是说，这种方法消除了分割线产生间断的问题。

尽管刚刚讨论的过程是用一个简单的例子说明的，但是处理更为复杂情况的方法是完全相同的，包括图 10.45(c)中显示的 3×3 对称结构元素的使用也是相同的。

10.5.3 分水岭分割算法

令 $M_1, M_2, \cdots, M_R$ 为表示图像 $g(x, y)$ 的局部最小值点的坐标的集合。如同在 10.5.1 节结尾说明的那样，这是一幅典型的梯度图像。令 $C(M_i)$ 为一个点的坐标的集合，这些点位于与局部最小值 M_i(回想无论哪一个汇水盆地内的点都组成一个连通分量)相联系的汇水盆地内。符号 min 和 max 代表 $g(x, y)$ 的最小值和最大值。最后，令 $T[n]$ 表示坐标 (s, t) 的集合，其中 $g(s, t) < n$，即：

$$T[n] = \{(s, t) \mid g(s, t) < n\} \tag{10.5.1}$$

在几何上，$T[n]$ 是 $g(x, y)$ 中的点的坐标集合，集合中的点均位于平面 $g(x, y) = n$ 的下方。

随着水位以整数量从 $n = \min + 1$ 到 $n = \max + 1$ 不断增加，图像中的地形会被水漫过。在水位漫过地形的过程中的每一阶段，算法都需要知道处在水位之下的点的数目。从概念上来说，假设 $T[n]$ 中的坐标处在 $g(x, y) = n$ 平面之下，并被“标记”为黑色，所有其他的坐标被标记为白色。然后，当我们在水位以任意增量 n 增加的时候，从上向下观察 xy 平面，会看到一幅二值图像。在图像中黑色点对应于函数中低于平面 $g(x, y) = n$ 的点。这种解释对于理解下面的讨论很有帮助。

令 $C_n(M_i)$ 表示汇水盆地中点的坐标的集合。这个盆地与在第 n 阶段被淹没的最小值有关。参考前一段的讨论，$C_n(M_i)$ 也可以被看做由下式给出的二值图像：

$$C_n(M_i) = C(M_i) \cap T[n] \tag{10.5.2}$$

换句话说，如果 $(x, y) \in C(M_i)$ 且 $(x, y) \in T[n]$，则在位置 (x, y) 有 $C_n(M_i) = 1$。否则 $C_n(M_i) = 0$。对于这个结果几何上的解释是很简单的。我们只需在水溢出的第 n 个阶段使用“与(AND)”算子将 $T[n]$ 中的二值图像分离出来即可。$T[n]$ 是与局部最小值 M_i 相联系的集合。

接下来，我们令 $C[n]$ 表示在第 n 个阶段汇水盆地被水淹没的部分的合集：

$$C[n] = \bigcup_{i=1}^{R} C_n(M_i) \tag{10.5.3}$$

然后令 $C[\max+1]$为所有汇水盆地的合集：

$$C[\max + 1] = \bigcup_{i=1}^{R} C(M_i) \tag{10.5.4}$$

可以看出(习题 10.29)处于 $C_n(M_i)$和 $T[n]$中的元素在算法执行期间是不会被替换的，而且这两个集合中的元素的数目与 n 保持同步增长。因此，$C[n-1]$是集合 $C[n]$的子集。根据式(10.5.2)和式(10.5.3)，$C[n]$是 $T[n]$的子集，所以，$C[n-1]$是 $T[n]$的子集。从这个结论我们得出重要的结果：$C[n-1]$中的每个连通分量都恰好是 $T[n]$的一个连通分量。

找寻分水线的算法开始时设定 $C[\min+1]=T[\min+1]$。然后算法进入递归调用，假设在第 n 步时，已经构造了 $C[n-1]$。根据 $C[n-1]$求得 $C[n]$的过程如下：令 Q 代表 $T[n]$中连通分量的集合。然后，对于每个连通分量 $q \in Q[n]$，有下列 3 种可能性：

(a) $q \cap C[n-1]$为空。
(b) $q \cap C[n-1]$包含 $C[n-1]$中的一个连通分量。
(c) $q \cap C[n-1]$包含 $C[n-1]$多于一个的连通分量。

根据 $C[n-1]$构造 $C[n]$取决于这 3 个条件。当遇到一个新的最小值时符合条件(a)，则将 q 并入 $C[n-1]$构成 $C[n]$。当 q 位于某些局部最小值构成的汇水盆地中时，符合条件(b)，此时将 q 合并入 $C[n-1]$构成 $C[n]$。当遇到全部或部分分离两个或更多汇水盆地的山脊线的时候，符合条件(c)。进一步的注水会导致不同盆地的水聚合在一起，从而使水位趋于一致。因此，必须在 q 内建立一座水坝(如果涉及多个盆地就要建立多座水坝)以阻止盆地内的水溢出。正如前一节中的解释，当用 3×3 个 1 的结构元素膨胀 $q \cap C[n-1]$并且需要将这种膨胀限制在 q 内时，一条一个像素宽度的水坝是能够构造出来的。

通过使用与 $g(x,y)$中存在的灰度级值相对应的 n 值，可以改善算法效率；根据 $g(x、y)$的直方图，可以确定这些值及其最小值和最大值。

例 10.18 分水岭分割算法的说明

分别考虑图 10.46(a)和(b)中显示的图像和它的梯度。应用刚才讨论的分水岭算法得到图 10.46(c)中显示的梯度图像的分水线(白色路径)。这些分割的边界叠加在原图上示于图 10.46(d)。如在本节中开始时所注意到的，分割边界具有那些被连接起来的路径的重要性质。

10.5.4 应用标记

直接以前一节中讨论的形式使用分水岭分割算法通常会由于噪声和其他诸如梯度的局部不规则性的影响造成过度分割。如图 10.47 所示，过度分割足以令应用算法得到的结果变得毫无用处。此时，过度分割意味着分割区域过多。一个较实际的解决方案是通过合并预处理步骤来限制允许存在的区域的数目，这些预处理步骤是为将附加知识应用于分割过程而设置的。

用于控制过度分割的方法是以标记的概念为基础的。一个标记是属于一幅图像的连通分量。我们有与重要对象相联系的内部标记，还有同背景相联系的外部标记。选择标记的典型过程包括两个主要步骤：(1)预处理；(2)定义一个所有标记必须满足的准则集合。为了对此进

行说明,再次考虑图 10.47(a)。导致图 10.47(b)中过度分割结果的一部分原因是大量隐含的最小值。由于这些区域的尺寸很小,所以这些最小值中有很多是不相关的细节。在前面的讨论中,已经不止一次地提到,将很小的细节对于图像的影响降至最低的有效方法是用一个平滑滤波器对图像进行过滤。在这种特殊情况下,这是一种合适的预处理方案。

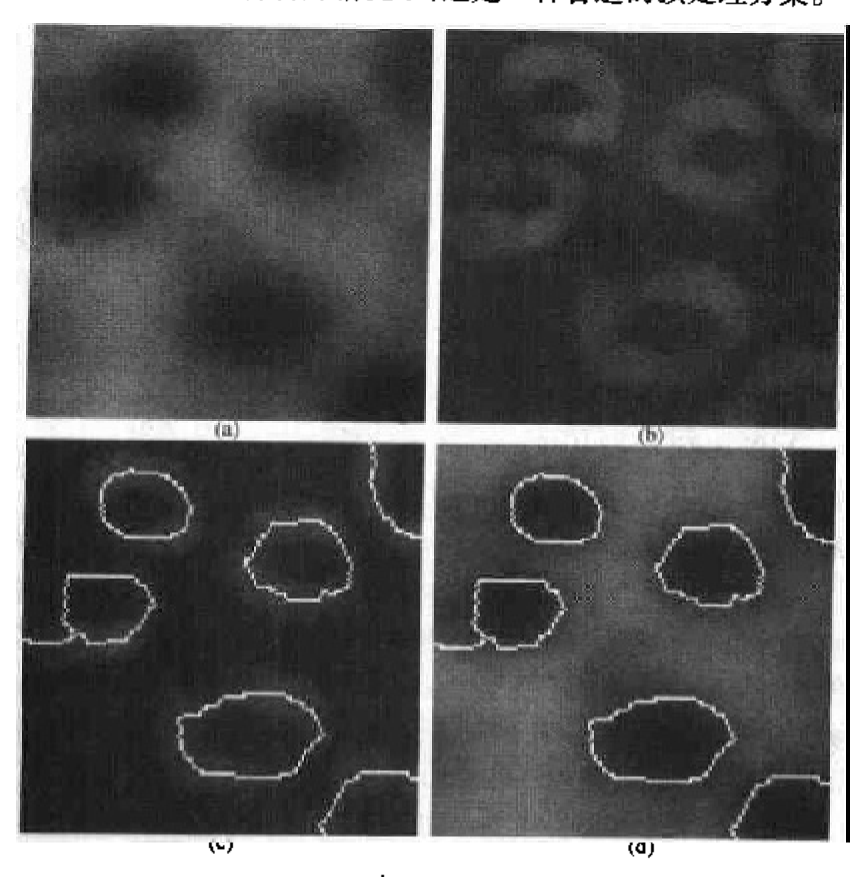


图 10.46 (a)带有斑点的图像,(b)梯度图像,(c)分水线,(d)叠加于原图中的分水线(由 CMM/Ecole des Mines de Paris 的 S. Beucher 博士提供)

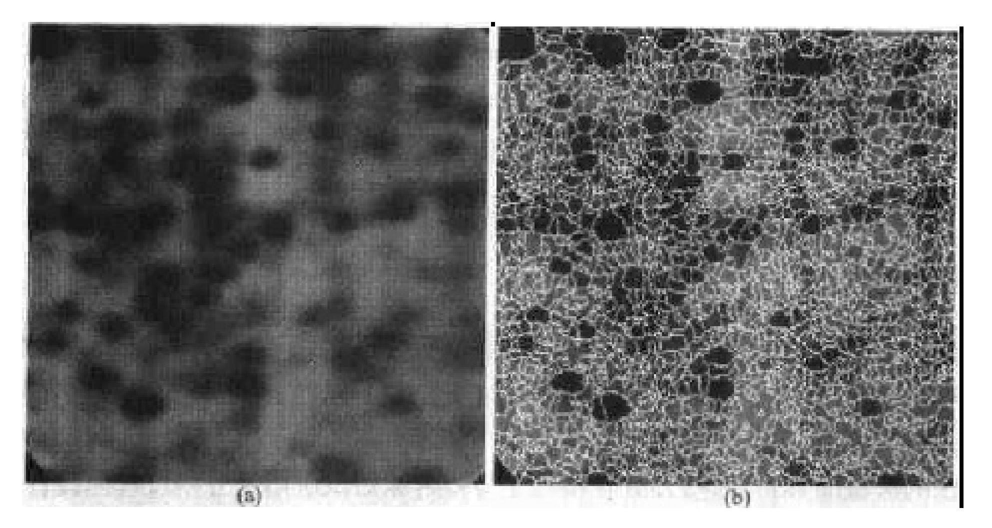


图 10.47 (a)电泳现象的图像,(b)对梯度图像使用分水岭分割算法得到的结果,过度分割现象很明显(由 CMM/Ecole des Mines de Paris 的 S. Beucher 博士提供)

假设在此时将内部标记定义为(1)被更高"海拔"点包围起来的区域;(2)区域中的点组成一个连通分量;并且(3)所有属于这个连通分量的点具有相同的灰度级值。在图像经过平滑处理之后,满足这些定义的内部标记以图10.48(a)中浅灰色、斑点状区域表示。下一步,对平滑处理后的图像使用分水岭算法,并限制这些内部标记只能是允许的局部最小值。图10.48(a)显示了得到的分水线。将这些分水线定义为外部标记。注意,沿着分水线的点是很好的背景候选点。因为它们经过相邻的标记之间的最高点。

图10.48(a)中显示的外部标记有效地将图像分割成不同区域。每个区域包含一个惟一的内部标记和部分背景。问题是因此变为将每个这样的区域一分为二:单一的对象和它的背景。我们对这个简单的问题能够应用多种在本章前面讨论过的分割技术。另一种简单的方法是对每个单独的区域使用分水岭分割算法。就是说,我们只求得平滑后的图像的梯度[如图10.46(b)],然后约束算法只对包含特定区域中标记的分水岭进行操作。使用这种方法得到的结果显示于图10.48(b)中。相对于图10.47(b)的改善之处是显而易见的。

标记的选择可以用基于灰度级值和连通性的简单的过程分类,如刚才所说明的那样,更复杂的描述涉及尺寸、形状、位置、相对距离、纹理内容等等(见第11章关于描绘子的部分)。关键是标记的使用给我们带来关于分割问题的某种带有先验性质的知识。应该提醒读者注意的是:人类经常以先验知识用日常的视觉辅助,进行各种分割和更高级的工作,最熟悉的例子就是在阅读上下文时所用到的方法。因此,分水岭分割方法提出了一种能有效使用这类知识的机制。这是这种方法的一个突出优点。

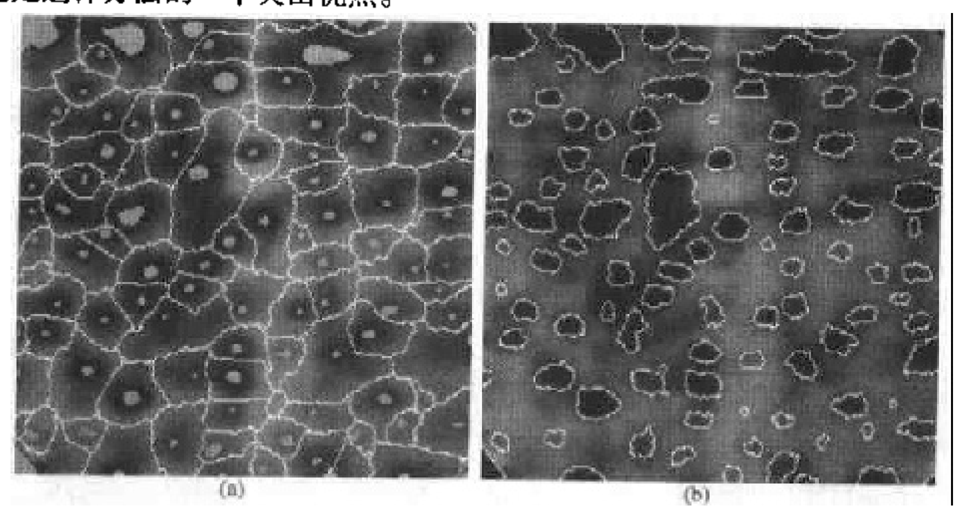


图10.48　(a)显示有内部标记(浅灰色区域)和外部标记(分水线)的图像,(b)分割的结果。注意其相对于图10.47(b)的改进之处(由CMM/EcoledesMinesdeParis的S. Beucher博士提供)

10.6　分割中运动的应用

运动是人类和动物使用的用于将重要的对象从不相关的背景细节中提取出来的强有力的提示。在成像应用领域,运动是由于感觉系统和视觉场景之间的相对位移而形成的,比如机器人应用、自主导航和动态视觉分析等等方面都是如此。在下面的各节中,我们将考虑运动在空间和频域分割中的运用。

10.6.1 空间技术

基本方法

分别检测两帧图像 $f(x,y,t_i)$和 $f(x,y,t_j)$之间在时间 t_i 和 t_j 时的变化的最简单的方法是将两幅图像逐个像素进行对比。这样做的过程是形成一幅差值图像。假设我们有一幅仅包含不动部分的参考图像。将这幅图像和随后生成的同一场景但包含一个运动对象的图像相比较,去掉两幅图像中的固定元素,余下的与非固定部分相对应的非零项就是两幅图像的差值。

两幅图像在时间 t_i 和 t_j 时的差值图像可以定义为:

$$d_{ij}(x,y)=\begin{cases}1 & |f(x,y,t_i)-f(x,y,t_j)|>T \\ 0 & \text{其他}\end{cases} \tag{10.6.1}$$

这里 T 是一个特定的门限。注意,如果在两幅图像间灰度差在其坐标上有相当的不同时,$d_{ij}(x,y)$具有值 1,差异程度取决于事先确定的门限 T。假设所有的图像具有同样的尺寸大小。最后,注意到式(10.6.1)中的坐标(x,y)的值跨越了这些图像的尺寸,所以差值图像 $d_{ij}(x,y)$与上述这些序列中的图像大小相同。

在动态图像处理过程中,$d_{ij}(x,y)$中所有值为 1 的像素被认为是对象运动的结果。这种方法只有在两幅图像都做了空间配准且在 T 值设定的范围内亮度相对较为稳定的情况下才是合适的。实际上,在 $d_{ij}(x,y)$中的 1 值项常常是由于噪声造成的。很典型的,在差异图像中,这些点是孤立的,去除它们的简单方法是在 $d_{ij}(x,y)$中构造 1 的 4 连通或 8 连通区域,并且忽略所有输入项少于预定数目的区域。尽管这种方法会忽略掉小的和/或慢速运动的对象,但这种方法改进了在差值图像中保留运动结果的几率。

差异积累

当试图从一图像序列中将运动的部分分离出来时,源自噪声的孤立项并不是一个无足轻重的问题。尽管通过门限处理的连通性分析可以减少这类项的数目,但这类过滤操作也会去掉前一节中提到的小的或慢速运动的对象。处理这一问题的方法是考虑几帧图像中同一像素的变化,这样要将"记忆"引入这一处理过程。其所包含的思想是忽略在序列帧上零星出现的变化并将其归于随机噪声的影响。

考虑一个图像帧序列 $f(x,y,t_1),f(x,y,t_2),\cdots,f(x,y,t_n)$,并令 $f(x,y,t_1)$为基准图像。一幅累积差异图像(ADI)是由将基准图像和图像序列中后续图像进行对比得到的。当基准图像和序列中图像之间在某个像素位置上出现一次差异就令一个计数器计数一次,这种计数器在积累差异图像的每个像素的位置上都有一个。因此,当第 k 帧图像与基准图像相比较时,积累差异图像中一个给定像素的输入项给出在此位置上对应的像素与基准图像中同一位置的像素间灰度级变化的次数。差异是设定好的,例如通过使用式(10.6.1)进行。

通常考虑三种类型的积累差异图像是有用的:绝对 ADI、正 ADI 和负 ADI。假设运动对象的灰度值大于背景的灰度值,这三种 ADI 可以被定义如下。令 $R(x,y)$表示基准图像,且为了简化符号,令 k 表示 t_k,则有 $f(x,y,k)=f(x,y,t_k)$。假设 $R(x,y)=f(x,y,1)$,然后记住作为计数的ADI值,并对所有 $k>1$,对(x,y)的所有相关的值定义如下:

$$A_k(x,y)=\begin{cases}A_{k-1}(x,y)+1 & |R(x,y)-f(x,y,k)|>T\\ A_{k-1}(x,y) & \text{其他}\end{cases} \tag{10.6.2}$$

$$P_k(x,y)=\begin{cases}P_{k-1}(x,y)+1 & [R(x,y)-f(x,y,k)]>T\\ P_{k-1}(x,y) & \text{其他}\end{cases} \tag{10.6.3}$$

和

$$N_k(x,y)=\begin{cases}N_{k-1}(x,y)+1 & [R(x,y)-f(x,y,k)]<-T\\ N_{k-1}(x,y) & \text{其他}\end{cases} \tag{10.6.4}$$

这里 $A_k(x,t)$,$P_k(x,y)$和 $N_k(x,y)$分别为遇到图像序列中的第 k 幅图像后的绝对 ADI、正 ADI 和负 ADI。

这些 ADI 是从所有计数值为零时开始这一点不难理解。同时要注意ADI如序列中的图像一样大小是相同的。如前面提到的那样,序列中的图像都被假设为同样大小。最后,我们注意到如果背景像素的灰度级大于运动对象的灰度级,则式(10.6.3)和式(10.6.4)中的不等式的顺序以及门限的符号是相反的。

例 10.19 绝对、正和负值积累差异图像的计算

图 10.49 显示了以 3 个亮度图像显示的 ADI,图中的矩形物体,大小为 75×50 像素,其中以每帧 $5\sqrt{2}$个像素的速度向东南方向运动。图像的大小为 256×256 像素。我们注意到下列几点:(1)正ADI的非零区域等于运动对象的大小;(2)正ADI的位置对应于基准帧中运动对象的位置;(3)当运动对象在基准帧中相对于同一对象完全被移动时,正ADI停止计数;(4)绝对 ADI 包括正的和负的 ADI;(5)运动对象的速度和方向可以由绝对的和负的 ADI 决定。

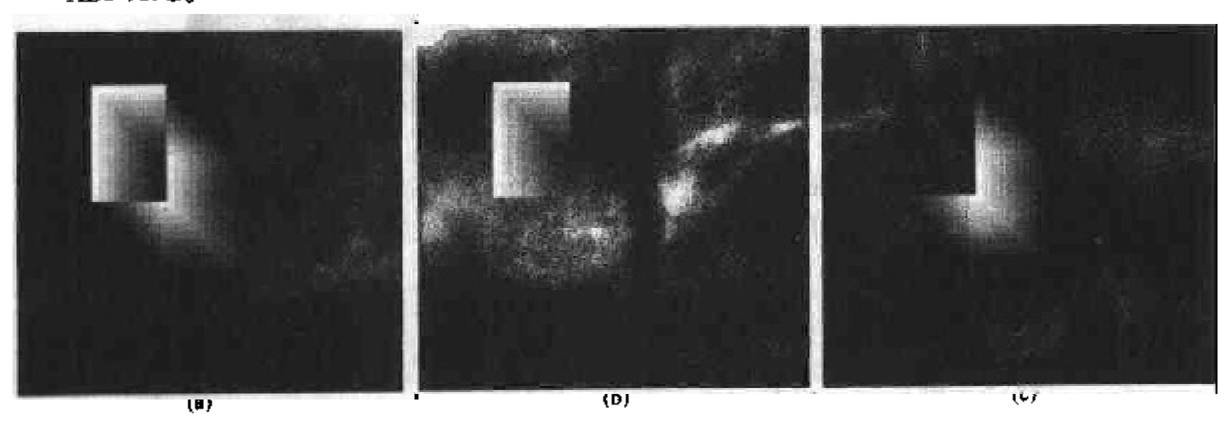


图 10.49 带有一个向东南方向移动的矩形目标的 ADI。(a)绝对 ADI,(b)正 ADI,(c)负 ADI

设置基准帧

前两节中讨论的技术成功的关键在于找到一幅可以与后续图像进行比较的基准图像。如上述指出的那样,在动态成像问题中两幅图像的差异倾向于消除图像中所有固定部分,而只留下对应于噪声和运动对象的图像元素。噪声问题可以通过使用前面提到的过滤方法或通过前面一节中提到的构造积累差异图像来解决。

实际上,用固定不动的元素构造一幅基准图像并不总是可行的,所以有必要根据一组包含

一个或多个运动对象的图像构造一个基准。特别是在描述变化频繁的场景或图像需要频繁更新的情况。生成基准图像的过程如下所述。考虑将图像序列中的第一幅图像作为基准图像。当一个非恒定的图像分量完全离开了它在基准帧中所处的位置时,当前帧中对应的背景可以被复制到在基准帧中由运动对象占据的位置。当所有的运动对象都离开了它们原先的位置时,就生成了只包含恒定分量的基准图像。正如前节所说明的那样,可以通过监控正的ADI中产生的变化完成对象的移动。

例 10.20　构造一个基准图像

图 10.50(a)和(b)显示了两幅关于一个交通十字路口的图像帧。第一幅图像被作为基准,第二幅在一段时间后描绘了同一场景。目的是去掉基准图像中主要的运动对象,生成一幅静态图像。尽管还有其他更小的运动对象,但主要的运动特征是十字路口上从左到右运动的汽车。为了说明,我们将注意力主要放在这个对象上。通过监视正的ADI的变化,可以如前所解释的那样确定运动对象的初始位置。一旦确定这个区域被此对象占据,可以通过剪影法将对象从图像中除去。通过观察图像序列中正的ADI停止变化的那一帧图像,可以从这幅图像中复制在初始帧中原来由运动对象占据的区域。然后这个区域就被贴到被挖去对象的图像中,从而恢复这个区域的背景。如果对所有运动的对象都进行这样的操作,结果正如在前两节中解释的那样,得到了一幅可以和后续帧相比,对于运动检测只包含静态图像部分的基准图像。此处,图 10.50(c)中显示了去掉向东运动的汽车后得到的结果。

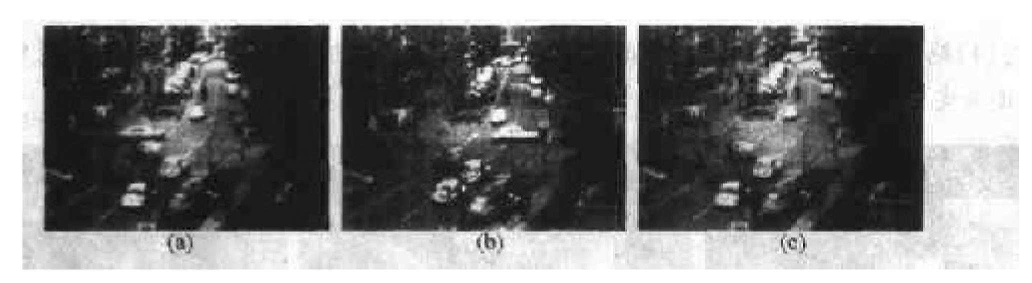


图 10.50　构造一幅静态基准图像,在图像序列中的(a)和(b)两帧图像,(c)从(a)中去掉向东运动的汽车并根据(b)中对应的区域恢复背景(Jain和Jain)

10.6.2　频域技术

本节中,我们考虑通过傅里叶变换确定对运动的估计的问题。考虑一个序列 $f(x,y,t)$,$t=0,1,\cdots,K-1$,由一台固定的照相机拍摄 K 帧大小为 $M\times N$ 的数字图像。通过假设所有帧有零亮度均一的背景开始讨论。一个例外是单一的,1 个像素的对象,它具有单位亮度并以恒定速度运动。假设对第一帧($t=0$),将图像平面投影到 x 轴上;即,将每一列像素的亮度相加。这样做会产生一个带有 M 个零值项的一维阵列,而对象的投影位置则是非零项。用 $\exp[j2\pi a_1 x\Delta t]$,$x=0,1,2,\cdots,M-1$,乘以阵列中的每个分量,包括在此帧所在的那一瞬间位于坐标(x',y')的对象在内,生成一个等于 $\exp[j2\pi a_1 x'\Delta t]$的和。在这个表示符号中,$a_1$ 是一个正整数,Δt 是不同帧之间的时间间隔。

假设在第二帧中($t=1$),对象运动到坐标$(x'+1,y')$;即,对象在平行于 x 轴的方向上移动了一个像素的位置。然后重复上一节中提到的投影过程,得到和 $\exp[j2\pi a_1(x'+1)\Delta t]$。如

果对象继续每一帧移动一个像素的位置，在任何整数的瞬间，结果是 $\exp[j2\pi a_1(x'+t)\Delta t]$，使用欧拉公式，可以用如下形式表示：

$$e^{j2\pi a_1(x'+t)\Delta t} = \cos[2\pi a_1(x'+t)\Delta t] + j\sin[2\pi a_1(x'+t)\Delta t] \tag{10.6.5}$$

$t=0,1,\cdots,K-1$。换句话说，这一过程得到一个频率为 a_1 的复正弦曲线。如果对象以每帧 v_1 个像素的速度运动(在 x 轴方向上)，正弦曲线的频率就是 v_1a_1。因为 t 在 0 到 $K-1$ 之间以整数增量变化，限制 a_1 只取整数值，使复正弦曲线的离散傅里叶变换具有了两个波峰(一个位于频率 v_1a_1 的位置上，另一个位于 $K-v_1a_1$ 的位置)。后一个波峰可能会被忽略掉。它是由于离散傅里叶变换的对称性而生成的，这一点在 4.6 节中进行了讨论。因此，在傅里叶频谱中找到一个波峰得到 v_1a_1。用 a_1 除这个量得到 v_1，它就是在 x 轴方向上的速度分量，这里假设帧频是已知的。同样可以通过这种方法得到 y 轴上的速度分量 v_2。

没有运动发生的帧序列产生同样的指数项，这些帧的傅里叶变换结果在频率为 0 处(一个单一的直流项)由单峰组成。所以，由于迄今为止讨论的都是线性操作，在静态背景中涉及一个或多个运动对象的一般情况下，对应于静态图像分量的直流的傅里叶变换结果为一个波峰，而在与对象的速度成比例的位置上有多个波峰。

上述这些概念可以总结如下。对一个有 K 幅大小为 $M\times N$ 的数字图像的序列，在任何值为整数的瞬时点上，图像加权投影于 x 轴上的和是：

$$g_x(t,a_1) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y,t)e^{j2\pi a_1 x\Delta t} \quad t = 0,1,\cdots,K-1 \tag{10.6.6}$$

同样，y 轴上的投影之和为：

$$g_y(t,a_2) = \sum_{y=0}^{N-1}\sum_{x=0}^{M-1} f(x,y,t)e^{j2\pi a_2 y\Delta t} \quad t = 0,1,\cdots,K-1 \tag{10.6.7}$$

这里，a_1 和 a_2 是正整数。

式(10.6.6)和式(10.6.7)的一维傅里叶变换分别为：

$$G_x(u_1,a_1) = \frac{1}{K}\sum_{t=0}^{K-1} g_x(t,a_1)e^{-j2\pi u_1 t/K} \quad u_1 = 0,1,\cdots,K-1 \tag{10.6.8}$$

和

$$G_y(u_2,a_2) = \frac{1}{K}\sum_{t=0}^{K-1} g_y(t,a_2)e^{-j2\pi u_2 t/K} \quad u_2 = 0,1,\cdots,K-1 \tag{10.6.9}$$

实际上，如 4.6 节中讨论的那样，这些变换的分量是应用 FFT 算法进行计算的。

频率-速度关系是：

$$u_1 = a_1v_1 \tag{10.6.10}$$

和

$$u_2 = a_2v_2 \tag{10.6.11}$$

在这个公式中，速度单位是用每总帧时间内的像素数来表示的。例如，$v_1=10$ 可以解释为在 K 幅图像帧中运动对象经过了 10 像素。对于匀速摄取的图像帧，真实的物理速度取决于帧频和像素之间的距离。因此，如果 $v_1=10$，$K=30$，帧频是每秒两幅图像，像素之间的距离是 0.5 m，则在 x 轴方向上的真实的物理速度是：

$$v_1 = (10\,\text{像素})(0.5\,\text{m/像素})(2\,\text{帧/s})/(30\,\text{帧}) = 1/3\,\text{m/s}$$

速度的 x 分量的符号通过计算下式得到：

$$S_{1x}=\left.\frac{\mathrm{d}^2\mathrm{Re}[g_x(t,a_1)]}{\mathrm{d}t^2}\right|_{t=n} \tag{10.6.12}$$

和

$$S_{2x}=\left.\frac{\mathrm{d}^2\mathrm{Im}[g_x(t,a_1)]}{\mathrm{d}t^2}\right|_{t=n} \tag{10.6.13}$$

因为 g_x 是正弦的,所以如果速度分量 v_1 是正的,则在时间 n 时,S_{1x}和 S_{2x}在任意点有相同的符号。相反,S_{1x}和 S_{2x}的符号相反说明速度分量 v_1 是负的。如果 S_{1x}或 S_{2x}为零,我们会考虑在时间上离得最近的下一个点 $t=n\pm\Delta t$。同样的结论可以应用于计算 v_2 的符号。

例 10.21　通过频率域检测小运动目标

图 10.51 通过图 10.54 说明了刚才得到的方法的有效性。图 10.51 显示了一组 32 帧的对地观测卫星(LANDSAT)图像中的一张,这组图像的基准图像中加入了白噪声图像。这个图像序列中叠加了一个运动目标,这个目标以 x 轴方向上每帧 0.5 个像素的速度和 y 轴上每帧 1 个像素的速度运动。目标在图 10.52 中用圆圈标出。它在一个小区域内(9 像素)具有高斯亮度分布,并且用肉眼不易辨别。令 $a_1=6$, $a_2=4$,计算式(10.6.8)和式(10.6.9)得到的结果分别显示在图 10.53 和图 10.54 中。图 10.53 中位于 $u_1=3$ 处的波峰根据式(10.6.10)得到 $v_1=0.5$。同样,图 10.54 中的波峰根据式(10.6.11)得到 $v_2=1.0$。

选择 a_1 和 a_2 的原则可以用图 10.53 和图 10.54 作为辅助进行解释。例如,假设令 $a_2=15$ 取代 $a_2=4$。此时,图 10.54 中的波峰将位于 $u_2=15$ 和 $u_2=17$。因为,$v_2=1.0$,这会导致一个严重混淆的结果。正如在 2.4.4 节中讨论过的,混淆是由于采样过疏造成的(由于 u 的范围由 K 决定,所以在当前的讨论中使用的帧太少)。因为 $u=av$,一种可能性是选择 a 作为最接近 $a=u_{max}/v_{max}$的整数,这里 u_{max}是确定 K 的混淆频率限制,而 v_{max}是对象的速度最大值。

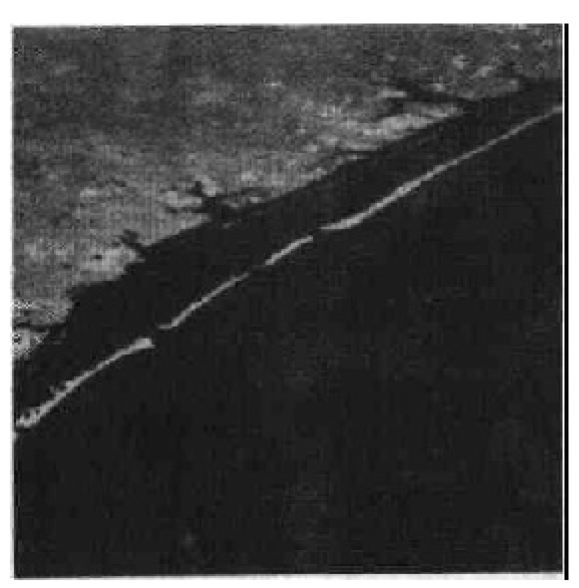


图 10.51　对地观测卫星图像中的一帧(Cowart, Snyder 和 Ruedger 提供)

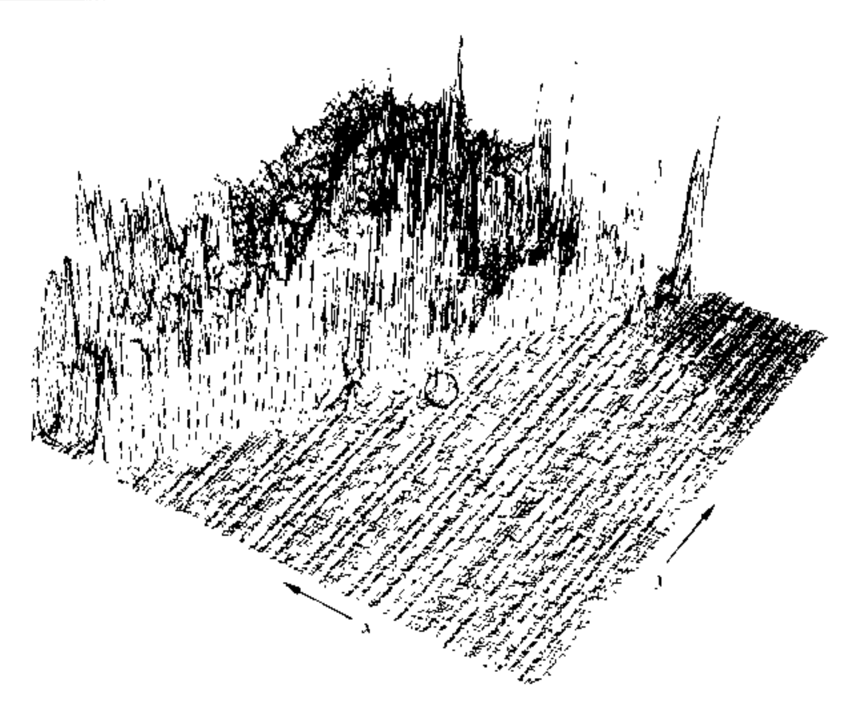


图 10.52 图 10.51 的图像亮度图。图中的目标用圆圈标出(Rajala, Riddle 和 Snyder 提供)

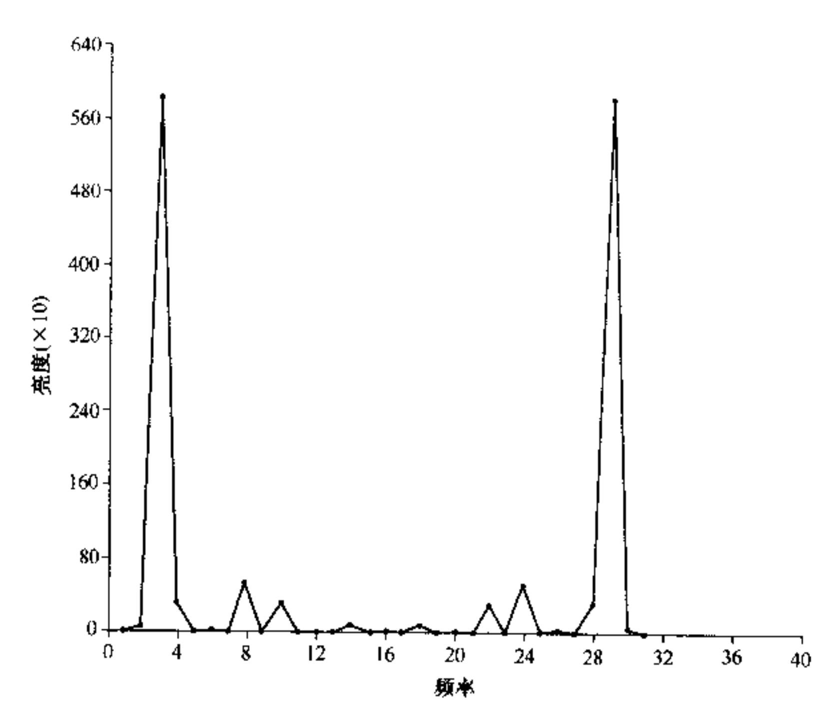


图 10.53 式(10.6.8)的频谱,显示了位于 $u_1 = 3$ 处的一个波峰(Rajala, Riddle 和 Snyder 提供)

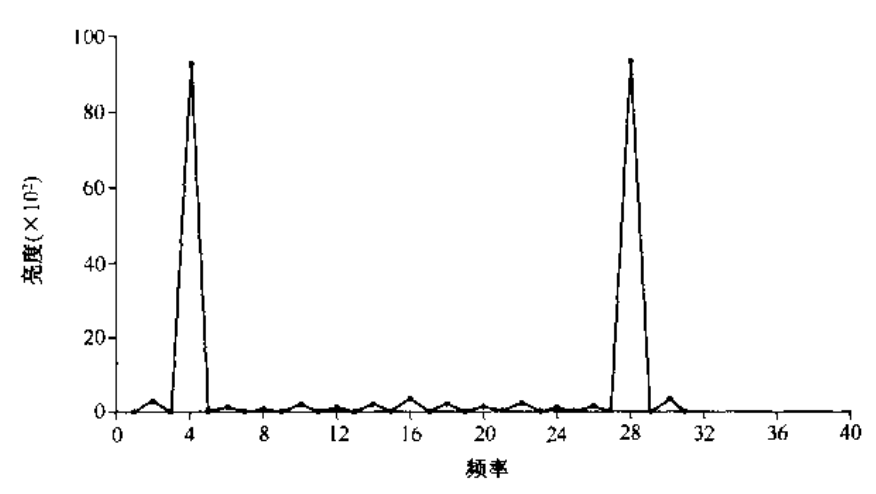


图 10.54 式(10.6.9)的频谱,显示位于 $u_2 = 4$ 处的一个波峰(Rajala, Riddle 和 Snyder 提供)

小结

图像分割在大多数自动图像模式识别和场景分析问题中是一个基本的预备性步骤。正如在前面各节中各种例子涉及的范围所说明的那样,选择一种分割技术而不选择另一种主要是由所面对问题的特点决定的。在本章中所讨论的各种方法,尽管并非十分的详尽,但确实是在实际应用中广泛使用的具有代表性的技术。下面的参考资料可以作为进一步研究这个主题的基础。

参考资料

由于分割在自动图像处理领域中所起的中心作用,因此,在大多数关于图像处理、图像分析和计算机视觉的著述中都涵盖了这一主题。下面提供一些书籍作为这一主题的阅读补充材料:Shapiro和Stockman[2001],Sonka等人[1999],Petrou和Bosdogianni[1999],以及Umbaugh[1998]。

使用模板检测灰度级间断具有很长的历史。这些年来人们提出了无数的模板,如Roberts[1965],Prewitt[1970],Kirsh[1971],Robinson[1976],Frei和Chen[1977]以及Canny[1986]。由Fram和Deutsch[1975]撰写的回顾性的文章包含了大量的模板以及对它们的效能的评价。模板效能存在的问题仍旧是相当重要的研究领域,特别是对边缘检测更是如此。这些在Qian和Huang[1996],Wang等人[1996],Heath等人[1997,1998]以及Ando[2000]的著述中均有例子。彩色图像的边缘检测在一些有多种意义的应用领域方面正逐渐普及开来。例如,Salinas, Abidi和Gonzalez[1996]的著述;Zugaj和Lattuati[1998]的著述;Mirmehdi和Petrou[2000]的著述;以及Plataniotis和Venetsanopoulos[2000]的著述。图像特征和模板效能之间的相互影响也是当前感兴趣的主题,如Ziou[2001]的著述中举的例子。对拉普拉斯算子的零交叉性质的表达是以Marr和Hildredth[1980]的论文以及Marr[1982]的著述为基础的。也可以参见Clark[1989]的论文,其中涉及使用零交叉算法进行边缘鉴别的内容(Clark的论文的部分修正内容在Piech[1990]中给出)。如 10.1 节中提到的,通过一个高斯型的拉普拉斯算子的零交叉方法是一种重要的方法。它相对于其他方法的效能仍是研究中相当活跃的主题(Gunn[1998,1999])。

霍夫变换(Hough[1962])结合了过去 10 年的研究成果提出了全局像素连接和曲线检测的

选择方法。许多以这种基本变换为基础而做的进一步推广的方法在本章中进行了介绍。它们都是多年以前就提出的方法。例如，Lo和Tsai[1995]提出了一种检测细线的方法，Cuil等人[1995,1997]解决了霍夫变换的快速实现问题和简单曲线的检测问题，Daul等人[1998]对检测椭圆弧的方法做了进一步的推广，Shapiro[1996]解决了对灰度级图像的霍夫变换的实现问题。在10.2.3节中提出的算法是以Martelli[1972,1976]的著述为基础的。关于试探性图的搜索，参见Nilsson[1980]，Umeyama[1988]和Sonka等人[1999]的相关著述。

如在10.3节的开始所提到的，门限技术享有特殊的地位。因为这类方法易于实现。所以有众多的著述在这个主题上着以相当多的文墨是并不奇怪的。从Sahoo等人[1988]和Lee等人[1990]的回顾性的文章中可以对这类文献所涉及的范围有一个较全面的评价。在10.3.2节关于门限处理中亮度的影响问题上，我们付出了相当大的努力。处理这类问题的方法在Perez和Gonzalez[1987]，Parker[1991]，Murase和Nayar[1994]，Bischsel[1998]以及Drew等人[1999]的相关著述中给予了说明。关于10.3.3节和10.3.4节的阅读材料参见Jain等人[1995]的著述。10.3.5节中讨论了Chow和Kaneko[1972]早期的工作。这些工作在说明基于门限的图像分割解决方案的关键内容方面仍被视为一种标准。实质上，10.3.6节中提到的材料也可以说具有相同的重要地位(由于White和Rohrer[1983]的著述)。这一节中我们将门限处理、梯度和拉普拉斯算子结合起来用于解决较为困难的分割问题。这两篇文章的时间与其在门限处理方面所做的工作相差了近20年(Cheriet等人[1998]，Sauvola和Pietikainen[2000])。比较它们得出的图像分割能力方面的基本相似点是非常有趣的事。也可以见Liang等人[2000]和Chan等人[2000]对于检测图像中的边界问题的替代方法的论述。在概念方面他们与Chwo和Kaneko在这方面所做的研究相似。

见Fu和Mui[1981]的一项面向区域的分割主题的综述。Haddon和Boyce[1990]以及Pavlidis和Liow[1990]所做的以分割为目的进行的关于将区域和边界信息合为整体的工作是在这方面最早进行的研究。Hojjatoleslami和Kittler[1998]提出的更进一步的区域生长方法也是很重要的。对于最近一段时期来面向区域的分割所涵盖的内容见Shapiro和Stockman[2001]以及Sonka等人[1999]的著述。

10.5节中对用分水岭方法分割进行的说明是一个有用的概念。这方面早期的分水岭分割的参考材料是Serra[1988]，Beucher[1990]以及Beucher和Meyer[1992]的有关著述。Baccar等人[1996]所写的文章对数据融合及形态学分水岭方法有所论述。十多年以来，在这个领域中，有关模式识别的空间性问题[2000]方面的进展是很明显的。这是在这些年来研究人员在这个主题上全身心投入的结果。如我们在10.5节的讨论中说明的那样，使用分水岭方法的关键问题之一是过度分割问题。Najmanand和Schmitt[1996]，Haris等人[1998]以及Bleau和Leon[2000]的文章对处理这类问题进行了说明。Bieniek和Moga[2000]提出了一种基于连通分量的分水岭分割算法。

10.6.1节中的材料是来自Jain, R.[1981]的著述。也可以参见Jain, Kasturi以及Schunck[1995]的著述。10.6.2节中的材料来自Rajala, Riddle和Snyder[1983]的著述。还可以参见Shariat和Price[1990]以及Cumani等人[1991]的文章。由Shapiro和Stockman[2001]写的以及Sonka等人[1999]写的书提供了关于运动估计方面的阅读材料。

习题

★**10.1** 有一幅包含水平的、垂直的、45°的和 - 45°直线的二值图像。给出一组 3×3 模板。这些模板可以用于检测这些直线中 1 个像素长度的间断。假设直线的灰度级是 1 并且背景的灰度级为 0。

10.2 提出一种技术用于检测二值图像的线段中长度从 1 到 L 个像素的裂缝。假设线是 1 个像素宽度。你的技术要以 8 邻域连通性分析为基础,而不能试图构造检测裂缝的模板。

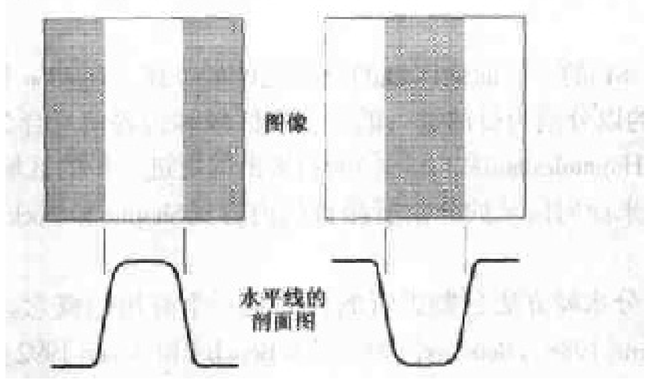
10.3 参考图10.4回答下列问题:

★(a) 图 10.4(b)中某些连接焊点和中心元素的线是单线,而其他一些是双线。解释一下为什么。

(b) 如何在图 10.4(c)中除去不属于 - 45°直线的部分?

10.4 考虑一幅穿过二值图像中部的水平亮度剖面图。二值图像包括一个垂直通过图像的中心的台阶状边缘。画出经过模糊处理后的图像的剖面图,处理的模板是大小为 $n \times n$,具有等于 $1/n^2$ 的系数的均值模板。为了简化,假设图像按比例绘制以便图像在边缘左边的灰度级为 0,右边的灰度级为 1。而且,假设模板的尺寸远小于图像尺寸,以便在水平亮度剖面图的中心附近不必关注图像的边界效应。

★**10.5** 假设使用图10.6中的边缘模型而不是边缘的斜面模型。写出每幅剖面图的梯度和拉普拉斯算子。



10.6 参考图10.8回答下列问题:

(a) 假设使用 Sobel 模板得到 G_x 和 G_y。说明此时使用式(10.1.4)和式(10.1.12)计算的梯度得出边缘在水平方向上和垂直方向上相同的结果。

(b) 说明这一结论对于 Prewitt 模板也是正确的。

★**10.7** 说明图10.8和图 10.9 的 Sobel 和Prewitt模板只对于水平、垂直及 ± 45°的边线能得出各向同性的结果。

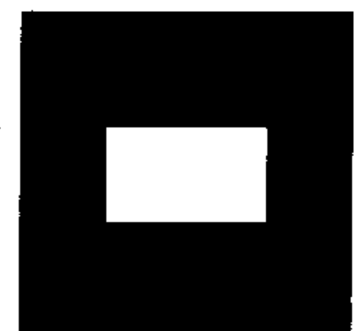
10.8 某些二维模板一遍处理得到的结果也可以通过一维模板两遍处理得到。例如,使用带有系数 1/9 的 3×3 平滑处理模板得到的结果,通过对图像使用一次模板[1 1 1]同样可以得到。得到的结果接着使用一次模板:

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

最后得到的结果乘以1/9。说明使用Sobel模板(图10.8)得到的结果可以通过使用一次形如[-1 0 1]的差分化模板(或其对应的正交阵),而后使用一次形如[1 2 1]的平滑处理模板(或其对应的正交阵)得到。

★10.9 所谓的罗盘梯度算子大小为3×3。它被用来计算8个方向上的梯度:E,NE,N,NW,W,SW,S和SE。写出系数值为0,1或-1时这8个算子的形式。确定每个模板的梯度方向,记住,梯度方向与边缘方向是垂直的。

10.10 在二值图像中心显示的矩形大小为 $m \times n$ 个像素。



(a) 使用式(10.1.12)给出的近似方法描绘此图像的梯度。假设 G_x 和 G_y 是通过使用Sobel算子得到的,说明梯度图像中所有相关的不同像素的值。

(b) 描绘根据式(10.1.5)计算的边缘方向的直方图。精确标记直方图中每个波峰的高度。

(c) 描绘式(10.1.14)给出的近似图像拉普拉斯算子。说明在拉普拉斯图像中所有相关的不同像素值。

10.11 参考式(10.1.17),

★**(a)** 说明拉普拉斯算子($\nabla^2 h$)的平均值为零。

(b) 证明所有用这个算子形成盘状结构的图像的平均值也是零。

(c) 在一般情况下,对于式(10.1.14)和式(10.1.15)所给出的拉普拉斯算子的近似,(b)还是正确的吗?请解释。

10.12 参考图10.15,

(a) 解释为什么图10.15(g)中的边缘构成闭合的轮廓线。

(b) 用于边缘定位的零交叉方法总是得到闭合轮廓线的结果吗?为你的答案找出原因。

★10.13 参考10.2.2节中的讨论。

(a) 解释为什么图10.20(b)中点1的霍夫映射是一条直线。

(b) 这是仅有的能产生这样的结果的点吗?

(c) 解释图10.20(d)中的邻接反射关系。

10.14 参考10.2.2节中的讨论。

(a) 根据直线的斜截式表示法 $y = ax + b$,寻找一种得到直线标准表示方法的一般过程。

(b) 找到直线 $y = -2x + 1$ 的标准表示方法。

10.15 图像分割技术的一个重要的应用领域是处理来自所谓的泡沫室现象的图像。这些图像是在高能物理实验中得到的。在实验中,一束已知性质的粒子直接打到已知核子的目标上。典型的现象包括输入轨迹,在撞击现象中由撞击点分散发出的粒子的二级轨迹。提出一种分割方法用于检测所有至少有100个像素并且下列6个方向上的所有的轨迹:±25°,±50°和±75°。在这6个方向上允许的估计误差为±5°。一条有效的轨迹必须不少于100个像素长并且间隙不多于3个,并且每个间隙不超过10个像素的长度。可以假设图像已经经过预处理以便形成二值图,且除了发出轨迹的撞击点以外所有轨迹均为1个像素宽。你的方案必须能够分辨原点不同但方向相同的轨迹(提示:以霍夫变换作为解决方案的基础)。

★**10.16** 参考图10.22和图 10.24。

(**a**) 将图 10.24 中的图给出的所有可能的边缘叠加到图 10.22 上。

(**b**) 计算最小开销路径的开销。

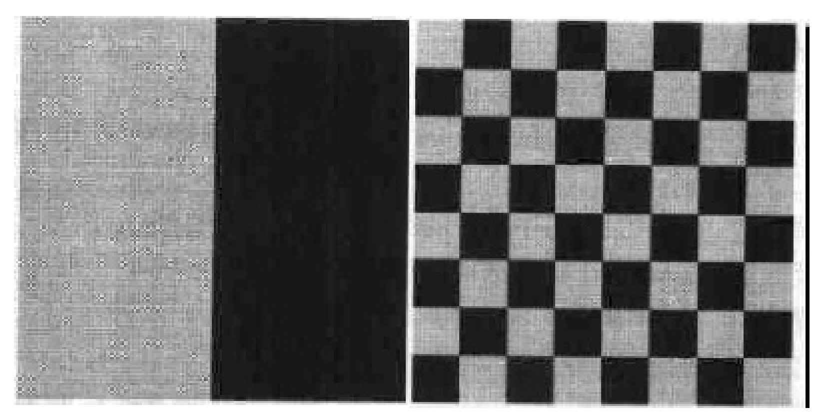
10.17 在显示的子图中找到对应于最小开销路径的边缘。括号中的数字是灰度级。外部的数字是空间坐标。假设边缘起始于第 1 列而结束于最后一列。

	1	2	3
1	• [2]	• [1]	• [0]
2	• [1]	• [1]	• [7]
3	• [6]	• [8]	• [2]

★**10.18** 下面显示的图像是很不相同的,但它们的直方图是相同的。假设每幅图像均用3×3的平滑处理模板进行模糊处理。

(**a**)模糊处理后的图像的直方图还是一样的吗?

(**b**)如果你的回答是不一样,则画出这两个直方图。



10.19 考虑一幅 $N\times N$ 大小的无噪声图像。此图像的前 $N/2$ 列具有灰度级 L_A,剩下的列有灰度级 L_B。这里 $L_B > L_A$。此图的直方图仅有两个高度相同波峰。一个位于 L_A 处另一个位于 L_B 处。根据灰度级将此图二等分是一种无意义的工作。可以通过使用设置于 L_A 和 L_B 之间的单一全局门限实现这种分割。然而,假设你用一个从左到右灰度级变化从 0 到 K 的楔形乘以此图像,此处 $K > L_B$。新的直方图是什么样子的? 对直方图的不同部分进行明确的标记。

10.20 参考10.3.3节中介绍的门限查找算法。假设在给定的问题中,直方图是双峰的并且模式的形状可以用形如 $A_1 e^{-(z-m_1)^2/2\sigma_1^2}$ 和 $A_2 e^{-(z-m_2)^2/2\sigma_2^2}$ 的高斯曲线进行拟合。假设 $m_1 < m_2$ 且当算法收敛时,下列给出的条件(根据这些曲线的参数)为真:

★(**a**)门限等于$(m_1 + m_2)/2$。

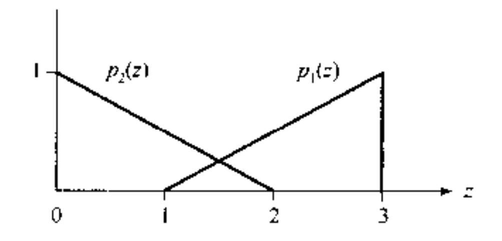
★(**b**)门限取小于 m_1 的值。

(**c**)门限在区间 $(m_1+m_2)/2<T<m_2$ 中取值。

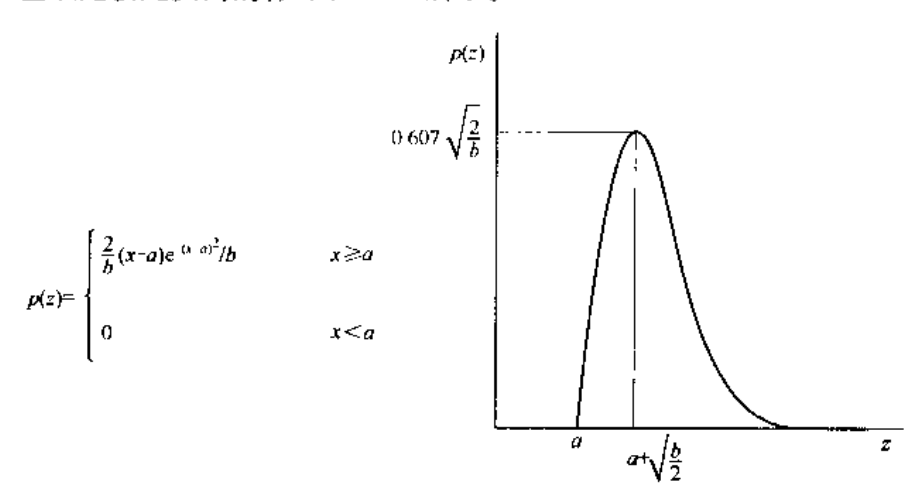
如果这些条件的任意一条不满足，请给出原因。

10.21 某一场景的照度 $i(x,y)$ 已知作为根据公式 $i(x,y)=A(1+e^{-B[(x-N/2)^2+(y-N/2)^2]})$ 得到的空间坐标的函数变化，并以标准单位进行测量。数字化的场景用 $f(x,y)$ 表示，场景的大小为 $N\times N$，$N=1\,000$。常量 $A=0.5$，$B=10^{-4}$。已知要得到 $f(x,y)$ 时，子图区域的任意两点之间亮度变化不超过 0.1 个单位，$f(x,y)$ 的任何大于 10×10 像素的子图可以被适当地分割。在这些条件下用于分割图像得到的子图的门限可以使用 10.3.3 节中介绍的算法求出。提出一种能够对这幅图像进行门限处理的自适应技术。不是所有的子图都大小相同，但在图像中的任何位置都需要使用可能的最大子图。

★**10.22** 假设图像的灰度级概率密度函数如图所示。这里 $p_1(z)$ 对应于对象，而 $p_2(z)$ 对应于背景。假设 $P_1=P_2$，找出对象像素和背景像素之间的最佳门限。



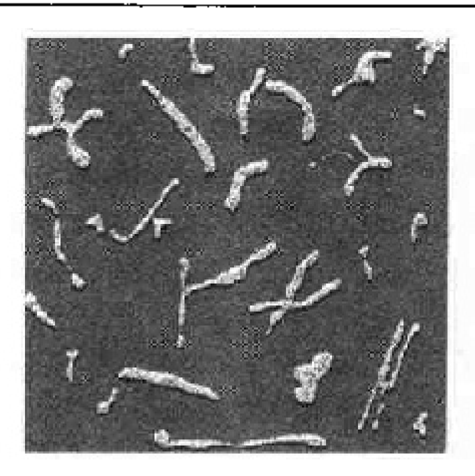
10.23 瑞利概率密度函数的函数形式和典型图形如下图所示。这个密度很适合于近似偏斜的直方图，如图 10.29 所示。使用瑞利密度提出一个与习题 10.22 一样的两类双峰问题。找出关于先验概率 P_1 和 P_2 的最佳门限和瑞利密度的参数。两个密度的长“尾巴”应该是彼此反向的，如图 10.29 所示。



★**10.24** 从式(10.3.10)开始推导出式(10.3.12)和式(10.3.13)。

10.25 从式(10.3.12)和式(10.3.13)开始推导式(10.3.14)。

★**10.26** 图像中背景像素的均值和标准差分别为 110 和 15。对象像素的均值和标准差分别为 200 和 40。提出一个用于从图像中分割对象的门限处理方案。清楚地阐述在处理这个问题时所做的假设。



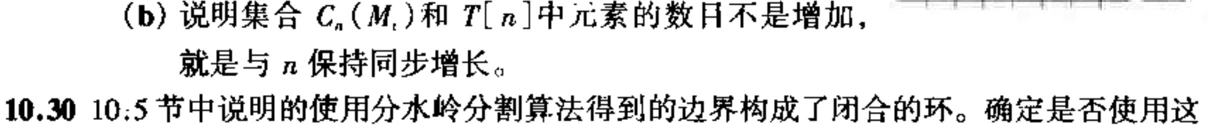
10.27 提出一种以区域生长为基础的分割习题 10.26 中图像的方法。清楚地阐述在处理这个问题时所做的假设。

10.28 通过使用10.4.3节中讨论的拆分和聚合过程分割所显示的图像。如果 R_i 中所有像素具有相同的灰度级，则令 $P(R_i)$ = TRUE。说明对应于你做的分割的四叉树。

10.29 参考 10.5.3 节中的讨论。

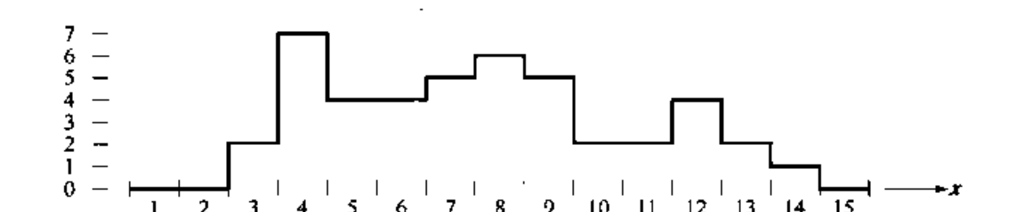
★**(a)** 说明在分水岭算法执行期间，集合 $C_n(M_i)$ 和 $T[n]$ 中的元素是不会被替换的。

(b) 说明集合 $C_n(M_i)$ 和 $T[n]$ 中元素的数目不是增加，就是与 n 保持同步增长。



10.30 10.5 节中说明的使用分水岭分割算法得到的边界构成了闭合的环。确定是否使用这种算法总会得到这样的闭合边界。

★**10.31** 对显示的一维灰度级交错部分逐步写出构建水坝的实现步骤。在每一步说明交错的部分，说明水位和水坝的形成过程。



10.32 如果我们在式(10.6.4)中使用 T 而不是 $-T$ 进行检验，则图 10.49(c)中的负ADI图像会是什么样子？

10.33 判断下列说法是否正确。对你的答案进行解释。

(a) 只要对象正在运动，绝对ADI的非零项的尺度就不断增大。

(b) 不管对象做何种运动，正ADI中的非零项总是占据相同的面积。

(c) 只要对象正在运动，负ADI的非零项的尺度就不断增大。

10.34 假设例10.21中将沿着 x 轴方向的运动设为 0。现在对象仅沿着 y 轴以每帧 1 个像素的速度运动 32 帧，然后瞬间转向，向相反方向运动 32 帧。在这些条件下，图 10.53 和图 10.54 会是什么样子？

10.35 飞行中的子弹的速度使用高速成像技术进行估计。选择的方法涉及电视照相机和可以照亮场景 K 秒的闪光灯。子弹 2.5 cm 长，1 cm 宽，速度范围在 750 ± 250 m/s。光学照相机生成的一幅图像中，子弹占据了分辨率为 256×256 的数字图像水平方向的 10%。

(**a**) 判断 K 的最大值以保证图像中由于运动造成的模糊不超过 1 个像素。

(**b**) 判断每秒必须得到的帧数的最小值，以便保证在子弹穿过照相机的视场时至少得到两幅子弹的完整图像。

(**c**) 提出一种自动从帧序列中提取子弹图像的分割程序。

(**d**) 提出一种自动判断子弹速度的方法。

第 11 章　表示与描述

在使用诸如第 10 章中谈到的那些方法将一幅图像分割成不同区域后，经常使用一种更适合于计算机进一步处理的形式，对得到的被分割的像素集进行表示和描述。基本上，表示一个区域包括两种选择：(1)可以用其外部特性来表示区域(如它的边界)，或(2)可以用其内部特性来表示区域(如组成区域的像素)。然而，选择一种表示方案仅仅是完成了将数据转换成可用于计算机处理的形式这个目的的一部分工作。下一步的任务是基于选好的表示方式，对区域进行描述。例如，可以用区域的边界表示这个区域，利用边界的特征对其进行描述，比如，边界的长度、连接边界上特殊点的直线方向和边界上凹陷的数目。

当关注的主要焦点集中于形状特性上时，可以选择外部表示法。而当主要的焦点集中于内部性质上时，可以选择内部表示法，比如颜色、纹理。有时，同时使用两种表示方法也是必须的。无论哪种情况，选择特征作为描述符号对于尺寸变化、平移和旋转都是很不灵活的。对于本章的大多数情况，我们所讨论的描述符号都能满足一种或更多的特性。

11.1　表示方法

第 10 章中讨论的分割技术生成原始数据，以形成沿着边界或处于区域内部的像素。尽管这些数据有时直接用于获得描述子(比如确定区域纹理时)，但标准的做法是，使用定好的方案将数据压缩为在描述子计算上相当有用的表示。在本节中，我们将讨论各种表示方法。

11.1.1　链码

链码用于表示由顺次连接的具有指定长度和方向的直线段组成的边界线。在典型的情况下，这种表示方法基于线段的 4 或 8 连接。每一段的方向使用数字编号方法进行编码，如图 11.1中所示。

获取或处理数字图像经常使用在 x 和 y 方向上大小相同的网格格式。所以，链码可以通过以顺时针方向沿着边界线，并且对连接每对像素的线段赋予一个方向生成。有两个原因使我们通常无法采用这种方法：(1)得到的链码往往太长，和(2)噪声或是边界线段的缺陷都会在边界上产生干扰。任何沿着边界的小干扰都会使编码发生变化，使其无法和边界形状相一致。

经常用来防止产生上述问题的方法是，选择更大间隔的网格对边界进行重新取样，如图 11.2(a)中所示。然后，由于网格线穿过边界线，则边界点就被指定为大网格的节点，这根据原始边界点最接近的节点为边界点的近似，如图 11.2(b)所示。使用这种方法得到的重新取样的边界可以用 4 或 8 链码表示，分别如图 11.2(c)和(d)所示。图 11.2(c)中的起始点(任意的)是在顶部左方的点，边界是图 11.2(b)的网格中容许的最短 4 或 8 通路。图 11.2(c)中的边界表达是链码 0033…01，图 11.2(d)是链码 0766…12。如预期的那样，编码表达方法的精确度依赖于取样网格的大小。

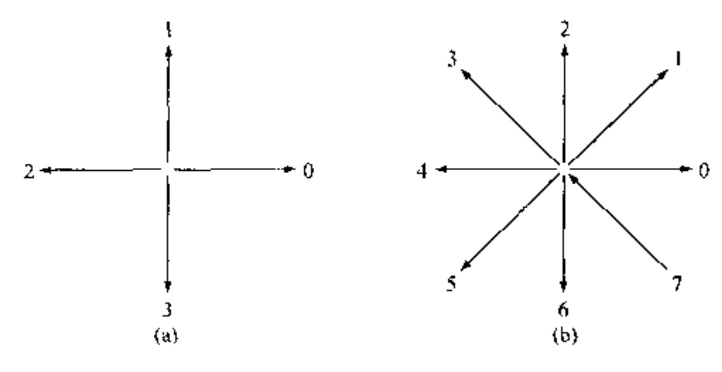


图 11.1 (a)4 向链码和(b)8 向链码的方向编号

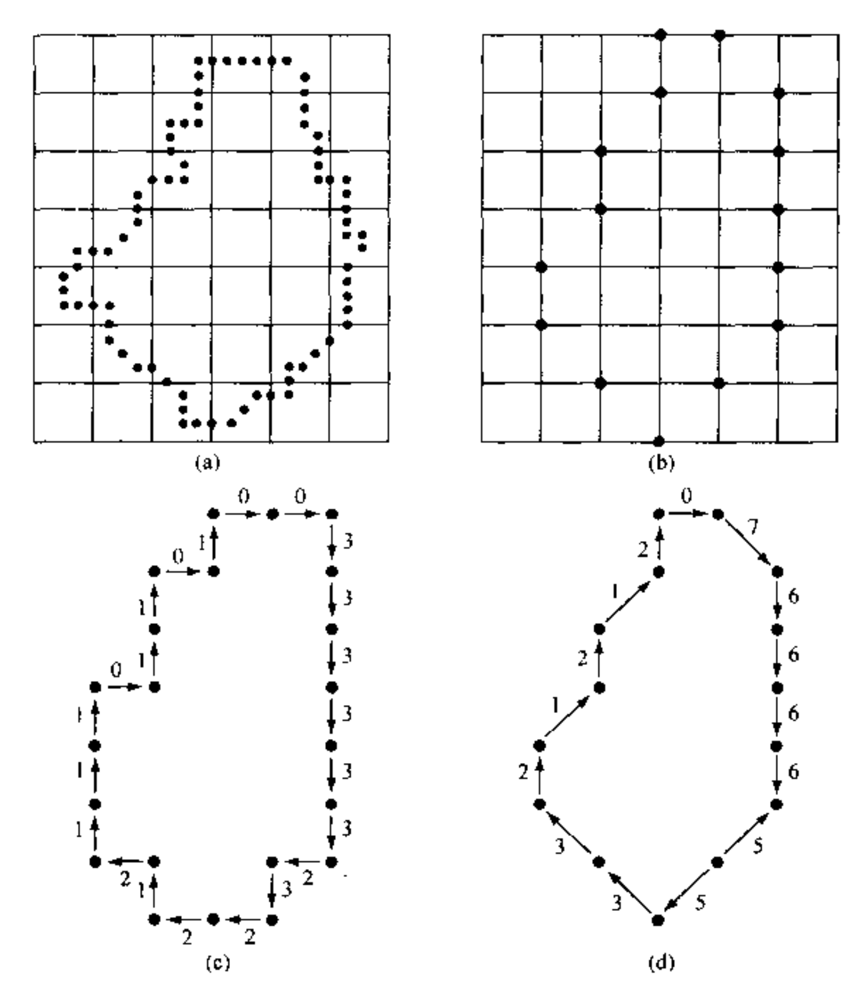


图 11.2 (a)叠加在数字化边界线上的重取样网格，(b)重取样的结果，(c)4向链码，(d)8向链码

边界的链码依赖起始点。然而，此编码可以通过简单的过程实现关于起始点的归一化：可以将链码看做方向编号的循环序列，并对起点进行重新定义以便得到的编号序列的整数值为最小值。也可以通过用链码的一次差分代替编码自身进行归一化，以便适应旋转变化。这一差分是通过计算相邻两个元素方向变化(逆时针方向)的数字得到的，这个变化将两个相邻元素的编码分开。例如，4 向链码 10103322 的一次差分是 3133030。如果把编码看做循环序列，

则差分的第一个元素是通过链的最后一个成员和第一个成员间的转移计算出来的。这里的结果是 33133030。尺寸的归一化可以通过改变取样网格的大小来实现。

这些归一化仅当边界自身无旋转和比例变化时才能保持其转换的准确性。实际上,这种情况很少见。例如,在两个不同方向上对同一对象数字化后,一般会生成不同的边界形状。其相异的程度与图像的分辨率成比例。这种影响可以通过在数字图像中选择像素间距比例大的链元素,或通过将用于重取样网格的方向定为被编码对象的主轴方向或定为此对象的本征轴线方向而得到减小。11.2.2 节和 11.4 节对此分别进行了阐述。

11.1.2 多边形近似

数字化边界可以用多边形进行任意精确性的近似。对一条闭合曲线,当多边形的边数等于边界线上的点数时,这种近似是准确的,此时,每对相邻点定义多边形的一条边。实际上,多边形近似的目的是,使用尽量少的多边形的边刻画边界图形的"本质"。通常,这不是一个一般的问题,解决这个问题的过程会很快转成一个耗时的循环搜索。然而,有几种在复杂性和处理需求方面较为适当的多边形近似技术还是很适合图像处理应用的。

最小周长多边形

先通过一种寻找最小周长多边形的方法论述多边形近似。这个过程用一个例子很好解释。假设用一系列彼此连接的单元将一条边界包住,如图 11.3(a)所示。这条由单元组成的环带使包围圈看起来像边界内外的两堵墙,而将对象边界想像成一条包含在墙中的橡皮圈。如果橡皮圈允许收缩,会形成图 11.3(b)的形状,生成一个有最小周长的多边形。这个多边形与单元组成的环带设置的几何图形相符。如果每个单元仅包含边界上的一个点,则在每个单元中,原来的边界和橡皮圈近似形之间的误差至多为$\sqrt{2}d$,这里 d 是不同点之间可能的最小距离(也就是,用于生成数字图像的取样网格中线条间的距离)。通过强制每个单元以它对应的像素为中心,这个误差可以减半。

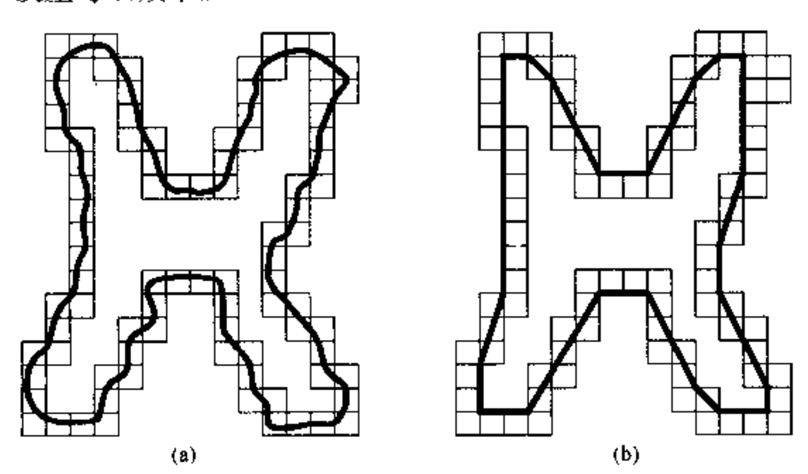


图 11.3 (a)被单元包围的对象边界,(b)最小周长多边形

聚合技术

基于平均误差或其他准则的聚合技术已经应用于多边形近似问题。一种方法是沿着边界

线寻找聚合点，直到适合聚合点的最小平方误差线超过一个预先设置的门限，这时就将点聚合。当这种情况出现时，直线的参量就被存储下来，误差设为 0，并且这个过程会不断重复下去，继续沿着边界线寻找，直到误差再一次超过门限值再聚合新的点。这一过程的最后，相邻线段的交点构成多边形的顶点。这种方法的一个主要难点在于，得到的近似图形的顶点并不总是与原来边界的拐点（比如拐角处）相一致，因为，新的线段只有超过误差门限的时候才开始画。例如，如果沿着一条长的直线追踪，而它出现了一个拐角，在超过门限之前，拐角上的一些点（取决于门限大小）会被丢弃。然而在聚合的同时进行拆分（下面会谈到）可以缓解这个难点。

拆分技术

边界线拆分的一种方法是，将一条线段不断地分割为两个部分，直到满足定好的某一标准。例如，可能出现这样的要求：即从边界线到某一直线的最大垂直距离不得超过预定门限，而这条直线要求连接此边界线的两个端点。如果这个条件满足，则距离此直线的最远点成为一个顶点，这样，将初始的线段再细分为两条子线段。这种方法在寻找突出的拐点时具有优势。对一条闭合边界线，最好的起始点是边界上的两个最远点。例如，图 11.4(a) 显示了一条对象的边界线，而图 11.4(b) 显示了对这条边界线的一次关于其最远点的再分割（实线）。标记为 c 的点是从顶部边界线段到直线 ab 的最远点（在垂直距离上）。同样，点 d 是从底部边界线段到直线 ab 的最远点。图 11.4(c) 显示了使用直线 ab 长度的 0.25 倍作为门限的拆分过程结果。

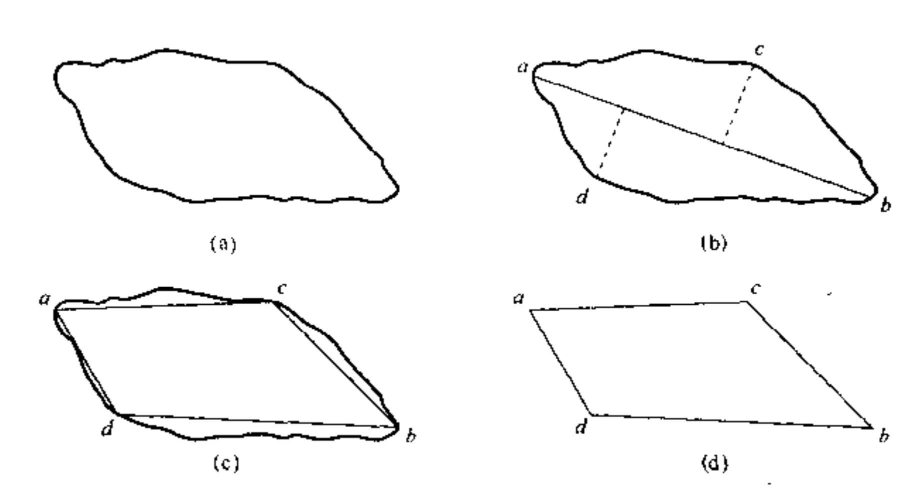


图 11.4　(a)原来的边界，(b)边界线被分割为基于端点的线段，(c)连接顶点，(d)得到的多边形

由于在新的边界线段上没有具有超过门限的垂直距离（相对于直线段）的点，分割过程终止于图 11.4(d) 所示的多边形。

11.1.3　标记图

一幅标记图是一种一维函数的边界表达方法，它可以用各种方法生成。最简单的方法之一就是将从质心到边界线的距离转化成一个角度函数，如图 11.5 所示。忽略标记图的生成，这种表达方式的基本思想是，假设一维函数表达会比原来的二维边界容易，因此使用一维函数

简化边界的表达。

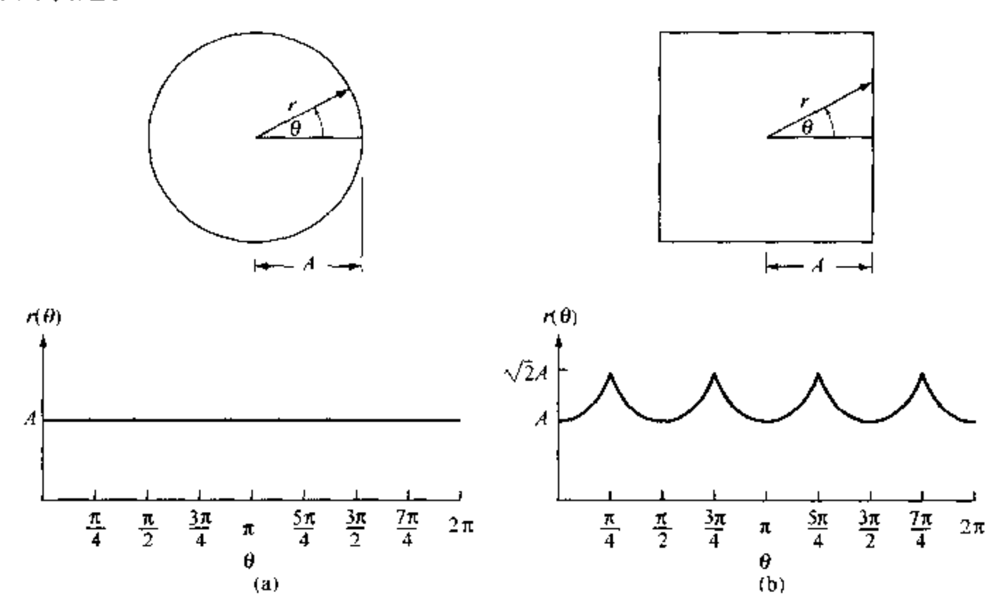


图 11.5 距离-角度的函数标记图。在(a)中 $r(\theta)$为常量。在(b)中,标记图由模式的重复出现构成,$r(\theta)=A\sec\theta \quad (0\leqslant\theta\leqslant\pi/4)$和$r(\theta)=A\csc\theta \quad (\pi/4<\theta\leqslant\pi/2)$

由刚刚论述的方法生成的标记图不会在转换过程中改变,但是标记图的生成依赖于旋转和比例缩放变换。通过寻找一种方法,选择相同的起点生成标记图而忽略图形的方向,可以实现旋转变换的归一化。实现这一目的的一种方法是,选择距离质心最远的点,如果这一点与我们关心的每个图形的旋转畸变无关,则选择这一点作为起点。另一种方法是在对象本征轴上(见 11.4 节),将离质心最远的一点定为起点。这种方法需要更多的计算,但更为严格,因为本征轴是由外形上的所有点决定的。而另一种方法是试图获得边界的链码,而后使用 11.1.1 节中谈到的方法,并假定得到的编码有足够的近似程度,使旋转不会影响到它的曲率。

基于两轴线缩放比例的一致性和以同一个值 θ 作为间隔进行取样的假设,形状尺寸的变化导致对应的标记图中幅值的变化。将这种结果进行归一化的一种方法是,对所有函数进行换算以便函数有相同的值域,比如[0,1]。这种方法的主要优点是简单。但是它也有潜在的严重缺陷,即对整个函数的缩放仅依赖于两个值:最小值和最大值。如果图形是带有噪声的,这种依赖性就可能成为从对象到对象的误差来源。一种更为严格的方法(但需要更繁重的计算)是依据标记图的变化对每个样本进行分割,并假设这种变化不为 0(如图11.5(a)中所示的情况)或小到造成计算困难。变化量的应用得到一个变化的缩放比例因子。这个因子与尺寸的变化是成反比的,并且像自动增益控制那样工作。无论使用什么方法,请记住基本思想是,在保存波形的基本形状时,消除其对尺寸的依赖性。

当然,距离-角度并不是生成标记图的惟一方法。例如,另一种方法是:穿过边界线,在边界线上的每个交点处画边界线在这一点的切线和基准线。得到的标记图尽管与 $r(\theta)$曲线很不同,但可以携带有关基本图形的特征信息。例如,曲线中的水平线段对应于沿着边界线的直线部分。因为在这里切线角度是常量。这种方法的变形是使用斜率密度函数作为标记图。这个函数只是简单的切线角度值的直方图。由于直方图是衡量值的密集程度的度量,斜率密度函数用恒定的切线角有利地反映了边界的各个部分(直的或近似直的线段),并且在角度变化

迅速的部分(拐角或其他突然的变化)呈现陡峭的凹陷。

11.1.4 边界线段

将边界线分段通常很有用。分段使边界的复杂性减小了,并且因此简化了描述过程。当边界线包含一个或更多携带形状信息的显著凹陷形状时,这种方法具有特殊的吸引力。此时,使用边界线闭合区域的凸壳就成为边界鲁棒性分解的有力工具。

如 9.5.4 节中定义的,一个任意集合 S 的凸壳 H 是包含 S 的最小凸集。集合之差 $H-S$ 称为集合 S 的凸形缺陷(Convex Deficiency) D。如何使用这些概念对边界进行有目的的分割,可参见图 11.6(a),图中显示了一个对象(集合 S)和它的凸形缺陷(阴影区)。区域边界可以通过跟踪 S 的轮廓并标记点(在该点向凸形缺陷之内或之外做转移)进行分割。图 11.6(b)显示了这一情况的结果。注意,在理论上,通常这种方案与区域的尺寸和方向无关。

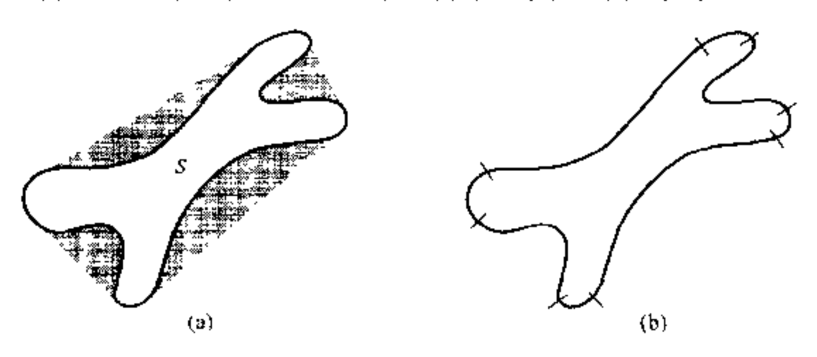


图 11.6　(a)区域 S 和它的凸形缺陷(阴影部分),(b)被分割的边界

实际上,数字边界由于数字化噪声和分割过程中的变形而易于变得不规则。这些影响通常导致凸形缺陷(它们且有较小的、通过边界随机扩散的无意义分量)。与其试图通过后处理找出这些不规则性,不如用通常的方法在分割之前先使边界变平滑。有很多方法可以做到这一点。一种方法是旋转边界并使用沿着边界的像素 k 个邻域的平均坐标代替这些像素的坐标值,这种方法可以处理小的不规则性,但它耗费时间并且难于控制。大的 k 值会导致过分的平滑,而小的 k 值也许使对边界某些部分进行的平滑处理变得不够充分。一种更为健全的技术是,在寻找到某个区域的图形缺陷之前先使用如 11.1.2 节中提到的多边形近似。我们关注的大部分数字化边界都是简单的多边形(没有自交叉的多边形)。Graham 和 Yao[1983]为寻找这种多边形的凸壳提出了一种算法。

凸壳概念和它的缺陷及区域的边界对于描述整个区域是同样有用处的。例如,描述一个区域可以基于它的面积和它的凸形缺陷的面积、在凸形缺陷中分量的数目、这些分量的相关位置,等等。9.5.4 节中探讨过寻找凸壳的形态学算法。参考本章结尾处含有其他表示法的引证材料。

11.1.5 骨架

一种表达平面区域结构形状的重要方法是把它简化成图形。这种简化可以通过使用一种细化(也称骨架化)算法得到区域的骨架来实现。细化过程在图像处理中起着核心作用并有着广泛的应用,从自动检测印刷电路板到对空气过滤器中的石棉纤维进行计数。已经在 9.5.7 节中讨论过使用形态学骨架化的基础理论。然而,正如在那一节中注意到的,讨论这一过程时

对如何保持彼此连接的骨架并未给出任何规定。在这里开发的算法将纠正这一问题。

一个区域的骨架可以用Blum[1967]提出的中轴转换方法(MAT)定义。边界为B的区域R的MAT如下所示。对R中的每个点p,找到它在B上最接近的邻点。如果p有多于一个这样的邻点,就认为p属于R的中轴(骨架)。“最接近”(和得到的MAT)这个概念取决于距离的定义(见2.5.3节)。图11.7显示了使用欧氏距离的一些例子。同样的结果可使用9.5.7节中提到的最大圆盘得到。

区域的MAT有一个基于所谓的“草原之火概念”的直觉定义。把图像区想像成一整块有着干草的平坦大草原,假如沿着草原的边界点火。火焰的前面会以同样的速度向区域中心前进。区域的MAT就是不止一处火焰的前沿在同一时刻所到达的点集合。

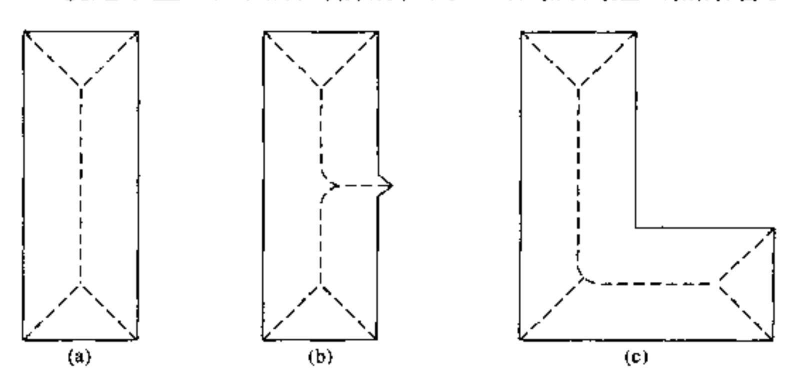


图11.7 3个简单区域的中轴(虚线)

尽管区域的MAT能生成较为合意的骨架,但直接实现这一定义需要大量计算。这其中包括计算区域的每个内部点到其边界点的距离。在试图产生一个区域中轴表达的同时,大量改进计算效率的算法也被提了出来。其中比较典型的算法是某些细化算法。这些算法在反复删除区域边界点时要受下列强制性约束:(1)不可删除端点,(2)不可破坏连通性和(3)不可造成对区域的过分侵蚀。

在这一节,我们提出一种应用于细化二值区域的算法。区域内的点假设具有1值,且背景点具有0值。这种方法对于给定区域的轮廓点逐次应用两个基本步骤,这里,给定区域是基于2.5.2节给出的定义,轮廓点是值为1的任意像素,并且在8个相邻点中至少有一个值为0的点。关于8个邻域的表示可以参照图11.8所示。第1步:如果下列条件满足,则标记的轮廓点p_1要删除。

$$\begin{aligned}&(\mathbf{a})\ 2 \leqslant N(p_1) \leqslant 6\\&(\mathbf{b})\ T(p_1)=1\\&(\mathbf{c})\ p_2 \cdot p_4 \cdot p_6 = 0\\&(\mathbf{d})\ p_4 \cdot p_6 \cdot p_8 = 0\end{aligned} \tag{11.1.1}$$

这里$N(p_1)$是p_1的非零相邻点的数目;即:

$$N(p_1) = p_2 + p_3 + \cdots + p_8 + p_9 \tag{11.1.2}$$

并且$T(p_1)$是以$p_2, p_3, \cdots, p_8, p_9, p_2$的次序轮转时,从0到1的变化次数。例如,图11.9中$N(p_1)=4, T(p_1)=3$。

p_9	p_2	p_3
p_8	p_1	p_4
p_7	p_6	p_5

图 11.8　用于细化算法的相邻关系排列

0	0	1
1	p_1	0
1	0	1

图 11.9　式(11.1.1)中条件(a)和(b)的图示说明，此时，$N(p_1)=4$且$T(p_1)=3$

在第 2 步，条件(a)和(b)仍相同，但条件(c)和(d)变为：

$$(\mathbf{c}')\ p_2\cdot p_4\cdot p_8=0$$
$$(\mathbf{d}')\ p_2\cdot p_6\cdot p_8=0 \tag{11.1.3}$$

在这种考虑下，对于二值区域的每个边界像素使用第 1 步。如果与条件(a)到(d)中一个或一个以上的条件不一致，则问题中的点值不改变。如果满足所有的条件，则点被标记以便删除。然而，要等到所有的边界点都经过处理后才能删除。之所以在所有边界点处理之后进行删除操作是为了防止在执行算法时改变数据结构。对所有的边界点执行了第 1 步后，将做了标记的点删除(改为 0 值)。然后，以同样的方式对所得结果数据执行第 2 步。

这样，细化算法的迭代包括：(1)执行步骤 1 对需要删除的边界点进行标记；(2)删除做了标记的点；(3)执行步骤 2，对剩下的要删除的边界点进行标记；(4)删除有标记的点。反复进行这一基本过程直到再没有可以删除的点，此时算法结束，生成了区域的骨架。

当边界点 p_1 仅有 1 个或 7 个 8 邻接点的值为 1 时，不满足条件(a)。只具有一个这样的点意味着点 p_1 是骨架某一笔画的端点，显然这个点是不能删除的。如果 p_1 有 7 个这样的邻接点，则删除 p_1 会对区域产生侵蚀。当对处在骨架一个像素宽的笔画上的点使用这种方法时，条件(b)是无法满足的。因此，这个条件防止在使用细化操作时断开骨架的线段。最小值的集合($p_4=0$ 或 $p_6=0$)或($p_2=0$ 和 $p_8=0$)同时满足条件(c)和(d)。因此，参考图 11.8 中的邻接排列关系，满足这些条件以及满足条件(a)和(b)的点是一个边界东边或南边的边界点或者边界西北角上的点。另一种情况是，p_1 不是骨架的一部分并将被删除。同样，下列最小值集合($p_2=0$ 或 $p_8=0$)或($p_4=0$ 和 $p_6=0$)同时满足条件(c′)和(d′)。这些值对应边界北边或西边的点或边界东南角上的点。注意，东北角上的点有 $p_2=0$ 和 $p_4=0$，既满足条件(c)和(d)，又满足条件(c′)和(d′)。对东南角上的点也是相同的，那里 $p_6=0$ 且 $p_8=0$。

例 11.1　一个区域的骨架

图 11.10 显示了一幅人类腿骨的分割图像并叠加了使用上述算法计算的区域骨架。在很大程度上，这个骨架在直观上是正确的。在骨骼“肩部”的右侧有一个双分支，人们第一眼看上去会想像有一个和左侧一样的单向分支。然而，请注意：右肩比左侧稍宽(在长的方向上)。那就是算法生成分支的原因。在骨架化算法中，这种不可预测的现象是常见的。

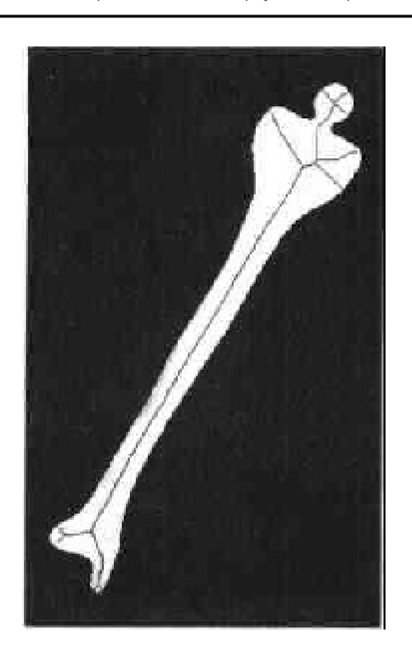


图 11.10 人腿骨和分层叠加显示的区域骨架

11.2 边界描绘子

这一节中我们考虑几种用于描绘区域边界的方法。在 11.3 节，我们关注区域描绘子。11.4 节和 11.5 节中的部分符号同时可用于边界和区域。

11.2.1 一些简单的描绘子

边界的长度是最简单的描绘子之一。一条边界上的像素数目可以大略表示其长度。对于在两个方向单位空间上定义的链码曲线，垂直和水平分量的个数加上$\sqrt{2}$倍的对角线分量的个数给出了其准确的长度。

边界 B 的直径定义为：

$$\mathrm{Diam}(B)=\max_{i,j}[D(p_i,p_j)] \tag{11.2.1}$$

这里 D 是距离的度量(见 2.5.3 节)，p_i 和 p_j 是边界上的点。直径的值和连接直径的两个端点的直线段(称为边界的长轴)的方向是表示边界的有用描绘子。边界线的短轴[①] 定义为与长轴垂直的直线。通过有两个轴线的边界 4 个外侧交点的方框长度完全包围了边界的轴线。刚刚描述的方框叫做基本矩形，长轴和短轴的比值叫做边界线的离心率。这也是一个有用的描绘子。

曲率定义为斜率的变化率。一般来讲，在数字化边界上找到某一点曲率的可靠量度是困难的，因为这种边界都较为“粗糙”。然而，有时使用相邻边界线段的斜率差作为线段交点处的曲率的描绘子，证明很有用。例如，像图 11.3(b)和图 11.4(d)中所示边界的顶点也可以很好地用于曲率描述。由于我们是顺时针方向沿着边界运动，当顶点 p 的斜率变化量为非负的时

① 不要将此处的长、短轴的定义同本征轴的定义混淆，后者在 11.4 节中定义。

候，称这一点属于凸线段；否则，称 p 属于凹线段。一点的曲率描述可以通过使用斜率变化的范围进一步精确化。例如，如果斜率的变化小于 10°，可认为它属于近似直线的线段。如果大于 90°，则属于拐点。然而，要注意，这些描绘子必须小心使用，因为相对于整条边界长度来说，这些描绘子的解释依赖于单独的线段。

11.2.2　形状数

如 11.1.1 节中所解释的那样，链码边界的首差取决于起点。这样的一条基于图 11.1(a) 所示的四向编码边界的形状数定义为最小数量级的首差。一个形状数的阶数 n 定义为其表达形式中的位数。再有，n 对于闭合边界是偶数，它的值限制了可能的不同形状的数目。图 11.11 显示了阶数为 4,6,8 的所有形状，与它们一起还显示了链码表达式、首差和对应的形状数。注意，首差是通过将链码作为循环序列计算后得到的，如 11.1.1 节所述。尽管链码的首差独立于旋转变化，但总体上说，编码边界依赖于网格的取向。将网格的取向归一化的一种途径是将链码的网格和前节定义的基本矩形结合起来。

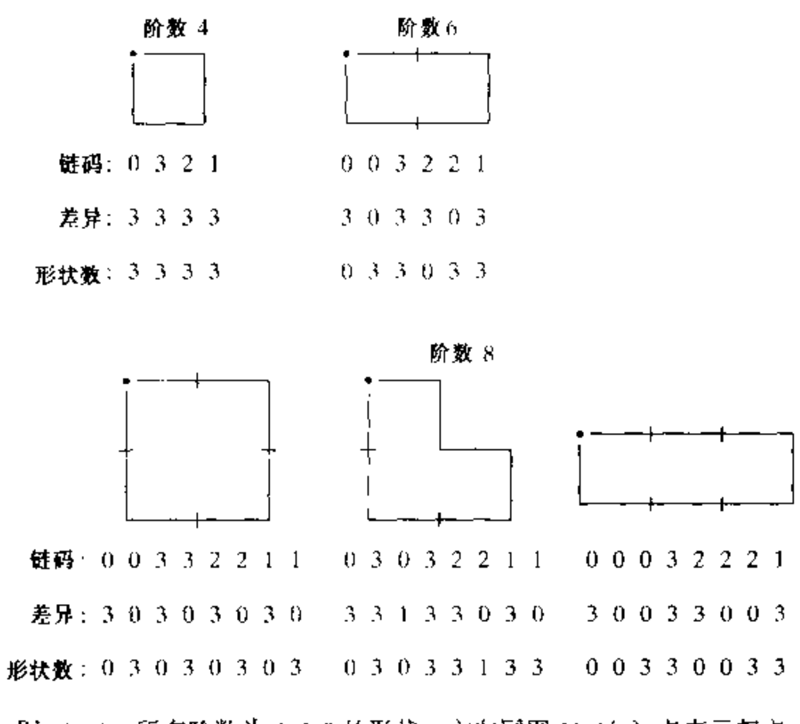
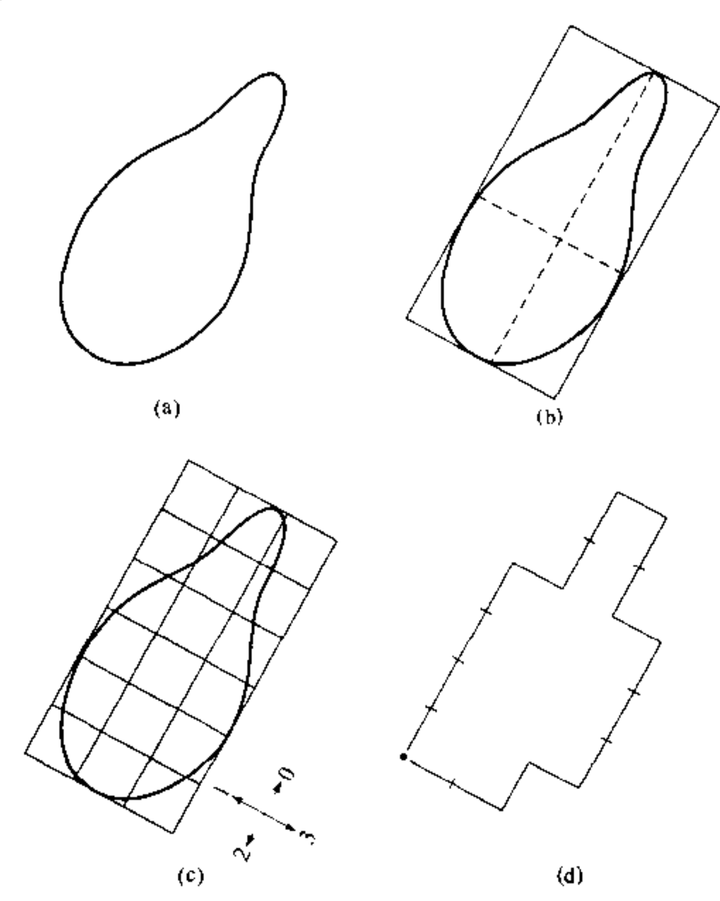


图 11.11　所有阶数为 4,6,8 的形状。方向同图 11.1(a)，点表示起点

实际上，对于一个想得到的形状数，我们要找到阶数为 n 的方框。这个方框所包含的曲线离心率（前节有定义）与基本方框的最近似，并使用这个新的方框设置网格尺寸。例如，如果 $n=12$，所有阶数为 12 的方框（即，周长长度为 12）是 2×4，3×3，1×5。如果对于给定的边界，2×4 的方框离心率最接近于此边界的基本方框离心率，就设置一个以此基本方框为中心的 2×4 网格，并应用 11.1.1 节的概述过程得到链码。形状数来自于链码的首差。尽管由于选定网格空间的方法使得到的形状数阶数一般等于 n，但有时相比这一间隔的边界减少，会使形状数的阶数大于 n。此时，我们指定一个阶数小于 n 的方框并且重复这一过程直到得到的形状数为阶数 n。

例 11.2 计算形状数

如图 11.12(a)所示,假设边界的阶数 $n=18$。为了得到这个阶数的形状数,需要遵循刚才提到的步骤。第 1 步是找到基本方框,如图 11.12(b)所示。最接近 18 阶的方框为 3×6 的方框,如图 11.12(c)所示的那样对基本方框进一步划分,这里链码的方向是依据选取的网格。最后一步是得到链码并用它的首差计算形状数,如图 11.12(d)所示。



链码: 0 0 0 0 3 0 0 3 2 2 3 2 2 2 1 2 1 1

差异: 3 0 0 0 3 1 0 3 3 0 1 3 0 0 3 1 3 0

形状数: 0 0 0 3 1 0 3 3 0 1 3 0 0 3 1 3 0 3

图 11.12 形状数的生成步骤

11.2.3 傅里叶描绘子

图 11.13 显示了一个 xy 平面内的 K-点数字边界。以任意点 (x_0, y_0) 为起点,坐标对 $(x_0, y_0), (x_1, y_1), (x_2, y_2), \cdots, (x_{K-1}, y_{K-1})$ 为逆时针方向沿着边界遇到的点。这些坐标可以用下列形式表示:$x(k)=x_k$ 和 $y(k)=y_k$。用这个定义,边界可以表示成坐标的序列 $s(k)=[x(k), y(k)]$, $k=0,1,2,\cdots,K-1$。再有,每对坐标对可以看做一个复数:

$$s(k)=x(k)+\mathrm{j}y(k) \tag{11.2.2}$$

对 $k=0,1,2,\cdots,K-1$。即对于复数序列,x 轴作为实轴,y 轴作为虚轴。尽管对序列进行了重新解释,但边界本身的性质并未改变。当然,这种表示方法的一大优点是:它将一个二维问题简化成一个一维问题。

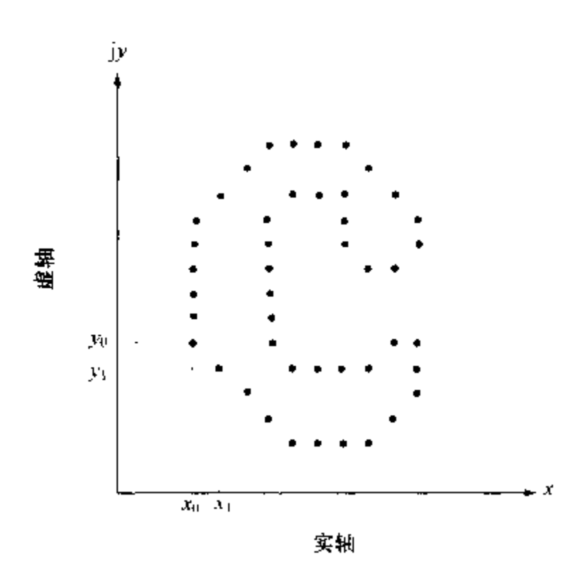


图 11.13　一条数字化边界和表示它的复数序列，点 (x_0,y_0) 和 (x_1,y_1)（任意的）是序列的前两个点

从 4.2.1 节得到，对离散 $s(k)$ 的傅里叶变换（DFT）为：

$$a(u)=\frac{1}{K}\sum_{k=0}^{K-1}s(k)e^{-j2\pi uk/K} \tag{11.2.3}$$

$u=0,1,2,\cdots,K-1$。复系数 $a(u)$ 称为边界的傅里叶描绘子。这些系数的反向傅里叶变换存于 $s(k)$。即：

$$s(K)=\sum_{u=0}^{K-1}a(u)e^{j2\pi uk/K} \tag{11.2.4}$$

对于 $k=0,1,2,\cdots,K-1$。然而，可以假设代替所有的傅里叶系数，只使用第一个 P 系数。这是式(11.2.4)设置 $a(u)=0,u>P-1$ 时的方程式。结果为 $s(k)$ 的近似值，如下所示：

$$\hat{s}(k)=\sum_{u=0}^{p-1}a(u)e^{j2\pi uk/K} \tag{11.2.5}$$

对于 $k=0,1,2,\cdots,K-1$。尽管只有 P 项用于计算 $s(k)$ 的每个元素 $\hat{s}(k)$，k 仍取从 0 到 $K-1$ 的值。即，在近似边界中存在同样数目的点，但重建每个点的时候并不使用同样多的项。回顾第 4 章中对傅里叶变换的讨论。高频元素能很好地解释细节，低频分量决定整体形状。因此，P 越小，边界细节失去得就越多。下面的例子给予了清楚的说明。

例 11.3　图示傅里叶描绘子

图 11.14 显示了一个包含 $K=64$ 个点的方形边界和对各种 P 值使用式(11.2.5)重建边界的结果。注意，重建边界前，P 值必须为 8，重建的边界比起圆形更像方形。接下来，注意直到 P 约为 56 时，拐角的点开始在序列中变得突出，符合拐角定义的变化才开始出现。最后注意，当 $P=61$ 时，曲线变直，此处几乎是一个原附加系数的精确复制。因此，一些低阶系数能够反映大体形状，而更多的高阶系数项是精确定义形状特征（比如拐角和直线）所需要的。从定义一个区域形状的过程中，低频和高频分量所起的作用来看，这个结果正是所期望的。

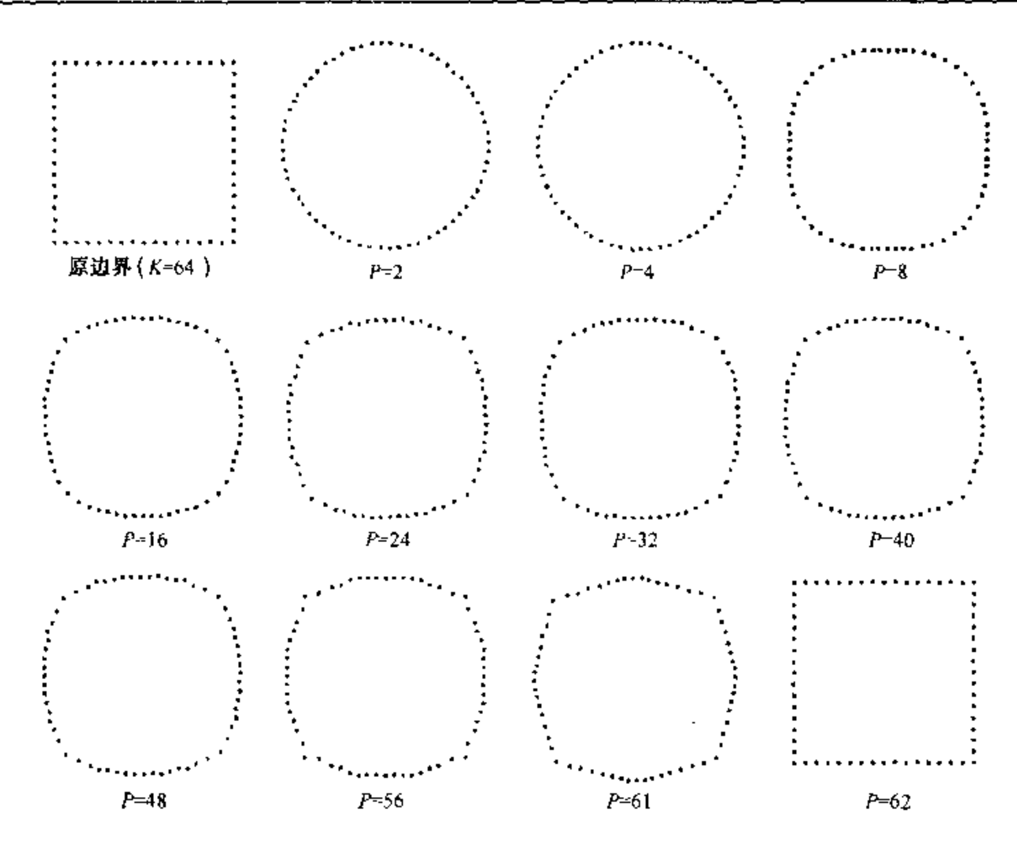


图 11.14 用傅里叶描绘子重建的例子。P 是重建边界使用的傅里叶系数的数目

如前面例子所做的示范,少数傅里叶描绘子能够反映边界的大略本质。这种性质很有用处,因为这些系数携带了形状信息。因此,如将在第 12 章中谈到的,这些描绘子可以作为区分不同形状边界的基础。

我们已经提到过多次:描绘子应该对平移、旋转和比例缩放变化不敏感。此处,描述边界的结果取决于经过处理的点的次序,所以,还有一个附加的限制条件就是,描绘子对起始点的位置不敏感。傅里叶描绘子对这些几何变化不敏感,但这些参数的变化可能关系到描绘子的简单变换。例如,考虑旋转变换并回顾基本数学分析中,关于一个点围绕复平面原点旋转 θ 角的变换。这种变换是通过乘以 $e^{j\theta}$因子实现的。对 $s(k)$的每个点进行这样的处理就实现了整个序列相对于原点的旋转变换。旋转的序列为 $s(k)e^{j\theta}$,其傅里叶描绘子为:

$$
\begin{aligned}
a_r(u) &= \frac{1}{K}\sum_{k=0}^{K-1} s(k)e^{j\theta}e^{-j2\pi uk/K} \\
&= a(u)e^{j\theta} \qquad (11.2.6)
\end{aligned}
$$

$u=0,1,2,\cdots,K-1$。因此,旋转变换只是通过乘以常数项 $e^{j\theta}$对所有系数给予相同的影响。

表 11.1 傅里叶描绘子的基本性质

变 换	边 界	傅里叶描绘子
原函数	$s(k)$	$a(u)$
旋转	$s_r(k)=s(k)e^{j\theta}$	$a_r(u)=a(u)e^{j\theta}$
平移	$s_t(k)=s(k)+\Delta_{xy}$	$a_t(u)=a(u)+\Delta_{xy}\delta(u)$
缩放	$s_s(k)=\alpha s(k)$	$a_s(u)=\alpha a(u)$
起点	$s_p(k)=s(k-k_0)$	$a_p(u)=a(u)e^{-j2\pi k_0 u/K}$

(注:$-j2\pi k_0 u/K$,$j\theta$ 均为 e 的次数。)

表 11.1 总结了边界序列 $s(k)$在执行旋转、平移、缩放和变换起点等变换过程中的傅里叶描绘子。符号 Δ_{xy} 定义为：$\Delta_{xy} = \Delta x + j\Delta y$，所以符号 $s_t(k) = s(k) + \Delta_{xy}$ 指明重定义的序列为：

$$s_t(k) = [x(k) + \Delta x] + j[y(k) + \Delta y] \tag{11.2.7}$$

换句话说，平移变换是对边界上的所有坐标加上一个位移常量。注意，除了 $u = 0$ 时变为冲激函数 $\delta(u)$[①]，变换过程对描绘子没有影响。最后，表达式 $s_p(k) = s(k - k_0)$表示重定义序列为：

$$s_p = x(k - k_0) + jy(k - k_0) \tag{11.2.8}$$

对序列的起点进行的变换仅是从 $k = 0$ 移到 $k = k_0$。从表 11.1 的最后一列乘以依赖于 u 的 $a(u)$这一点来看，起点的变化以不同的方式(但是是已知的)影响所有描绘子。

11.2.4 统计矩

边界线段的形状(和特征波形)可以通过简单的统计矩进行定量的描述，如均值、方差和高阶矩。要了解如何实现这一点，可以参考图 11.15(a)和(b)，前者显示了边界的线段，后者显示了以任意变量 r 的一维函数$g(r)$描述的线段。这个函数通过连接线段的两个端点并将线段旋转至水平方向得到。点的坐标也旋转相同的角度。

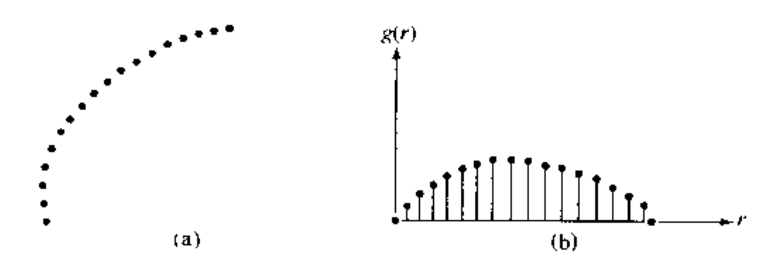


图 11.15　(a)边界线段，(b)一维函数表示法

将 g 的振幅看做离散随机变量 v 并形成关于 $p(v_i)$的直方图，$i = 0,1,2,\cdots,A-1$，这里 A 是分割振幅数值标尺的离散振幅增量。请记住，$p(v_i)$是生成值 v_i 的概率估计值，按照式(3.3.18)，v 的第 n 个估计值的均值为：

$$\mu_n(v) = \sum_{i=0}^{A-1}(v_i - m)^n p(v_i) \tag{11.2.9}$$

这里，

$$m = \sum_{i=0}^{A-1} v_i p(v_i) \tag{11.2.10}$$

m 是 v 的均值或平均值，μ_2 是它的方差。一般来讲，几乎不需要用一阶矩来区分形状明显不同的图形。

一种替代方法是将 $g(r)$归一化为单位面积下的函数并把它做成直方图。换句话说，是将 $g(r_i)$作为产生值 r_i 的概率。此时，将 r 作为随机变量，则矩为：

---

① 回顾第 4 章常量的傅里叶变换是一个位于原点的脉冲，而其他处均为 0。

$$\mu_n(r)=\sum_{i=0}^{K-1}(r_i-m)^n g(r_i) \tag{11.2.11}$$

这里,

$$m=\sum_{i=0}^{K-1}r_i g(r_i) \tag{11.2.12}$$

在这个式子中,K 是边界上点的数目,$\mu_n(r)$与形状函数 $g(r)$有直接关系。例如,二阶矩 $\mu_2(r)$用来衡量 r 的均值曲线分布,三阶矩 $\mu_3(r)$衡量以均值作为参考的对称性。

基本上,我们实现的是将描述工作简化为描述一维函数。尽管矩法显然是使用最为普遍的方法,但它们并不是实现这一目的的惟一描绘子。例如,另一种方法可以是计算一维离散傅里叶变换,得到它的谱,使用谱的前 q 个分量描述$g(r)$。矩法优于其他技术的优点是,实现起来简单并且带有对边界形状的"有形"解释。从图 11.15 也可以看出这种方法对于旋转变换的不敏感性。如果需要,通过缩放 g 和 r 的范围可以实现尺寸归一化。

11.3 区域描绘子

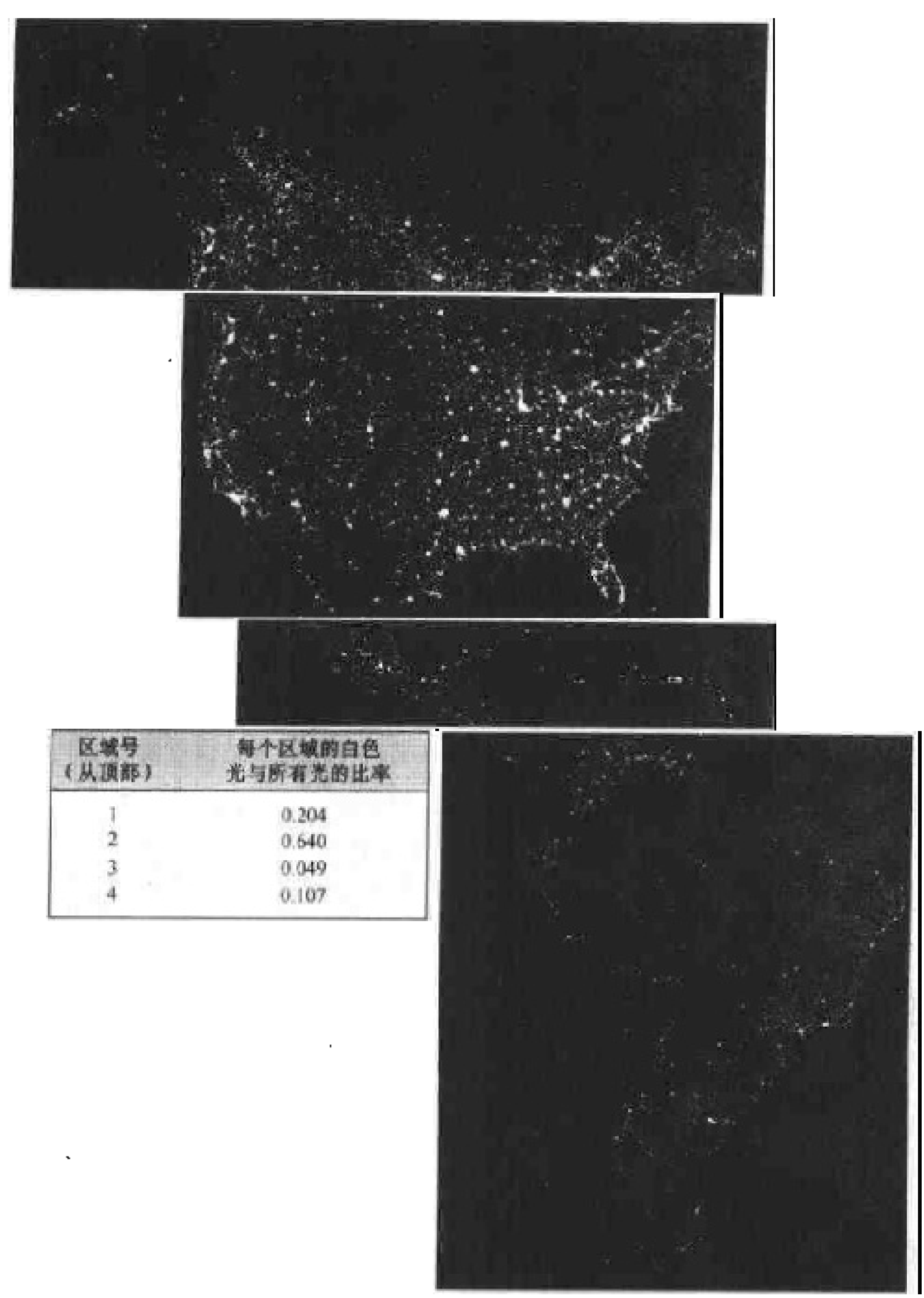
在这一节中,考虑描述图像区域的不同途径。要记住边界和区域描绘子结合使用是实践中普遍的做法。

11.3.1 某些简单的描绘子

一个区域的面积定义为区域中像素的数目。区域的周长是它的边界长度。尽管面积和周长有时用做描绘子,但它们主要在所关注的区域大小不变的情况下使用。这两个描绘子更为常用的地方是衡量一个区域的致密性,定义为"(周长)^2/面积"。致密性是无量纲的量(因此对均匀标度的变化不敏感)并且圆盘形区域的致密性是最小的。除了在数字区域的旋转变换时引入的误差之外,致密性对于方向性也是不敏感的。其他简单的用做区域描绘子的量包括灰度的均值和中值、最小和最大灰度级值及大于和小于均值的像素数。

例 11.4 使用面积计算从图像中提取信息

甚至像归一化面积这样简单的区域描绘子,在从图像中提取信息方面也是很有用处的。例如,图 11.16 显示了卫星拍摄的关于美洲的红外线图像。如在 1.3.4 节中讨论的细节一样,像这样的图片能提供区域内定居人数的总量。收集这些图像的传感器有能力检测可见物体和近红外线发射源,比如灯光、火和闪光。图像旁边的表格显示(从顶部到底部的区域)在所有 4 个区域中白色(光)所占面积与所有发光区域面积的比率。例如,像这样的简单测量可以得到消耗电能地区的相关估计值。通过将其归一化可以从数据中提取出有关大陆块的每个地区和人口数等信息。



区域号（从顶部）	每个区域的白色光与所有光的比率
1	0.204
2	0.640
3	0.049
4	0.107

图 11.16　夜间美洲的红外图(由 NOAA 提供)

11.3.2　拓扑描绘子

拓扑特性对于图像平面区域的整体描述是很有用处的。简单来说,拓扑学是研究一种在图像没有撕裂和连接的情况下(有时这被称为橡皮伸展变形),不受任何变形影响的图形性质。例如,图 11.17 显示了一个有两个孔的区域。如果一个拓扑描绘子由区域内孔洞数来定义,这种特性明显不受伸展和旋转变换的影响。然而,一般来讲,在区域发生分裂或聚合时,孔的数目会发生改变。注意,由于伸展影响距离量,因此拓扑特性也不依赖于距离概念和任何隐含地基于距离度量概念的性质。

另一个对区域描述有用处的拓扑特性是连通分量的数目。一个区域的连通分量在 2.5.2 节定义。图 11.18 显示了一个有 3 个连通分量的区域(见 9.5.3 节关于计算连通分量的算法)。

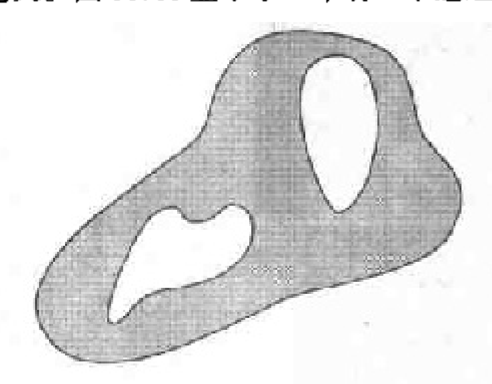




图 11.17 有两个孔的区域

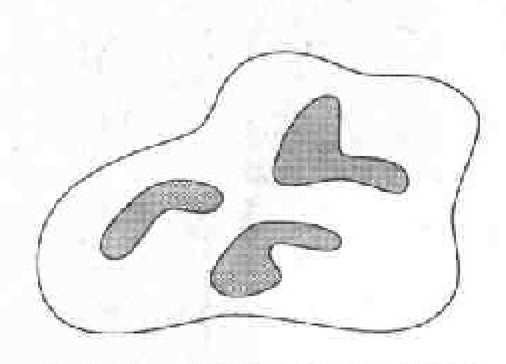


图 11.18 一个有 3 个连通分量的区域

图形中孔的数目 H 和连通分量 C 可以用于定义欧拉数 E:

$$E = C - H \tag{11.3.1}$$

欧拉数也是一种拓扑特性。如图 11.19 中显示的区域有分别等于 0 和 -1 的欧拉数,因为"A"有一个连通分量和一个孔,而"B"有一个连通分量和两个孔。

由直线段表示的区域(称为拓扑网络)用欧拉数解释会相当简单。图 11.20 显示了一个拓扑网络。将一个这样的网络进行内部区域分类,分成面和孔通常是很重要的。V 代表顶点数,Q 代表边数,F 代表面数,给出的下列关系称为欧拉公式:

$$V - Q + F = C - H \tag{11.3.2}$$

由式(11.3.1)看出它等于欧拉数:

$$V - Q + F = C - H = E \tag{11.3.3}$$

图 11.20 中显示的网络有 7 个顶点,11 条边,2 个面,1 个连通区域和 3 个孔;因此,欧拉数为 -2:

$$7 - 11 + 2 = 1 - 3 = -2$$

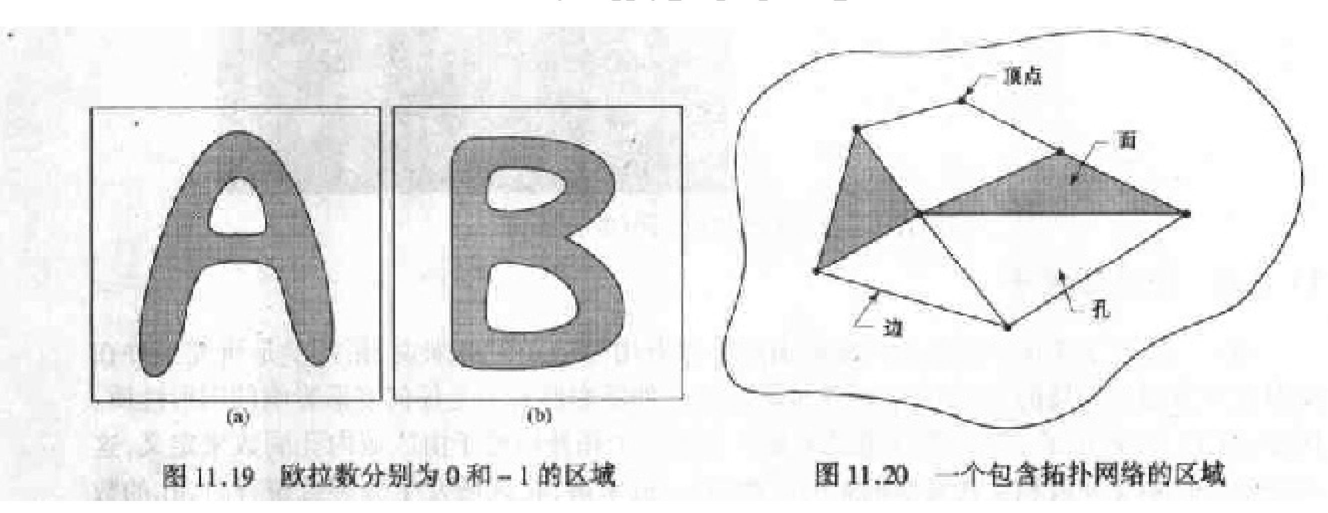


图 11.19 欧拉数分别为 0 和 -1 的区域

图 11.20 一个包含拓扑网络的区域

拓扑描绘子提供了一个附加特征,在描绘一幅场景的区域特性时非常有用。

例 11.5　用连通分量在分割后的图像中提取最大特征

图 11.21(a)显示了一幅 512×512 大小，8 比特的华盛顿特区图像，该图像由 NASA 陆地(LANDSAT)卫星拍摄。这幅特别的图像是在近红外波段拍摄的(细节见图 1.10)。假设只使用这幅图分割河流(不使用多频谱图像可以简化任务)。由于河流在图像中是相当暗且分布均匀的区域，设置门限值显然是一件试探性的工作。在使河流变为一连通区域之前，用可能的最高门限值为图像设置门限所得的结果如图 11.21(b)所示。门限值需要人工选择，这一点说明在这种情况下，仅靠河流本身的图像而没有经过门限值处理的其他区域的图像，是不可能将河流图像分离出来的。这个例子的目的就是，用图示说明连通分量如何用于"完成"图像分割。

图 11.21(b)中的图像有 1 591 个连通分量(使用 8 连通得到)。其欧拉数为 1 552，从这两点可以推断出有 39 个孔。图 11.21(c)显示了带有最大元素数目(8 479)的连通分量。这就是希望得到的结果，我们知道这个结果不可能由其自身图像分割。注意，所得结果是非常清晰的。如果要完成测量，如河流每条支流的长度，可以使用连通分量的骨架[图 11.21(d)]。换句话说，骨架中每个分量的长度近似接近图中河流的支流所显示的长度。

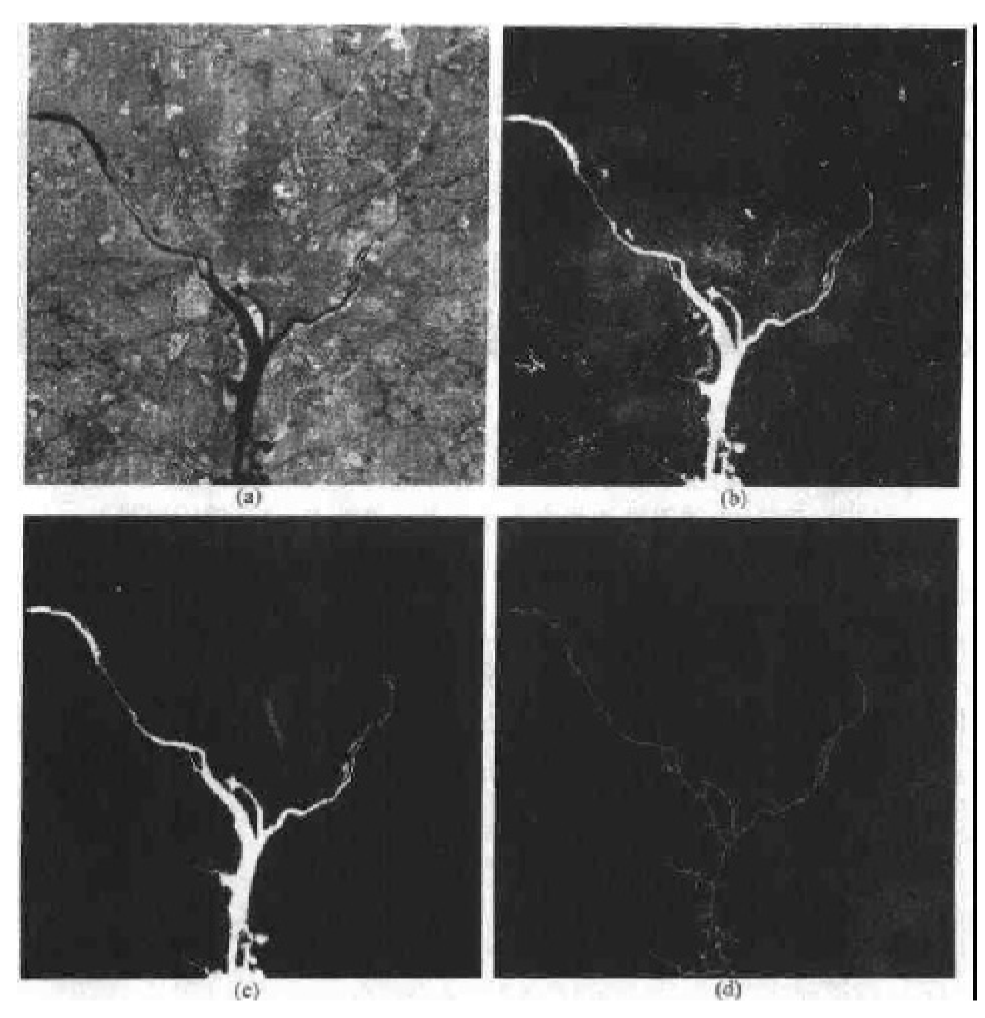


图 11.21　(a)华盛顿特区的红外图像，(b)经门限值处理的图像，(c)图(b)的最大连通分量，(d)图(c)的骨架

11.3.3 纹理

区域描绘的一种重要方法是将区域的纹理量化。尽管对纹理没有正式的定义,但在直觉上,这种描绘子提供了对诸如平滑度、粗糙度和规律性等特性的度量(图 11.22 中显示了一些样例)。在图像处理中用于描述区域纹理的 3 种主要方法是统计方法、结构化方法和频谱方法。统计方法指诸如平滑、粗糙、粒状等纹理的特征描述。结构化技术处理图像元的排列,诸如基于均匀空间分布的平行线纹理描述。频谱技术基于傅里叶频谱特性,主要用于通过识别频谱中高能量的窄波峰寻找图像中的整体周期性。

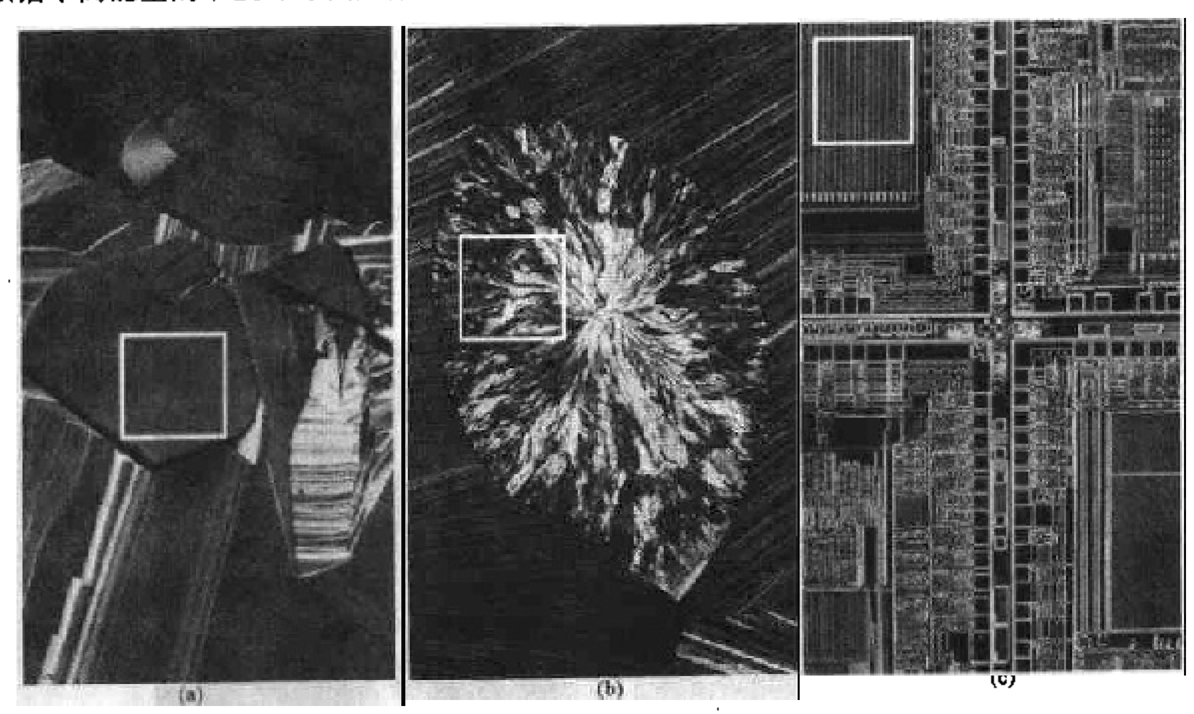


图 11.22 从左到右白色方块标记平滑、粗糙、有规则的纹理。这些光学显微镜下的图像分别是一块超导体、人类的胆固醇和一个微处理器(由佛罗里达州立大学的 Michael W. Davidson 博士提供)

统计方法

描述纹理的最简单方法之一是使用一幅图像或区域灰度级直方图的统计矩。令 z 为一个代表灰度级的随机变量,并令 $p(z_i), i=0,1,2,\cdots,L-1$,为对应的直方图,这里 L 是可区分的灰度级数目。由式(3.3.18),关于 z 的均值的第 n 阶矩为:

$$\mu_n(z) = \sum_{i=0}^{L-1}(z_i - m)^n p(z_i) \tag{11.3.4}$$

这里 m 是 z 的均值(平均灰度级):

$$m = \sum_{i=0}^{L-1} z_i p(z_i) \tag{11.3.5}$$

注意式(11.3.4)中,$\mu_0 = 1, \mu_1 = 0$。二阶矩[方差 $\sigma^2(z) = \mu_2(z)$]在纹理描述中特别重要。它

是表示灰度级对比度的量度，可以用于建立有关平滑度的描绘子。例如，量度：

$$R = 1 - \frac{1}{1 + \sigma^2(z)} \tag{11.3.6}$$

对于等强度区域为 0(此处方差为 0)，对于大的 $\sigma^2(z)$ 值接近于 1。因为方差值对于大值的灰度级图像是趋向于增大的，比如在 0 到 255 的范围内，所以将方差归一化后转换到区间[0,1]中，再用于式(11.3.6)是一个好主意。这可以通过在式(11.3.6)中用 $(L-1)^2$ 除 $\sigma^2(z)$ 实现。标准差 $\sigma(z)$ 也经常作为纹理的量度使用，因为标准差的值对于很多人来说更为直观。

三阶矩，

$$\mu_3(z) = \sum_{i=0}^{L-1} (z_i - m)^3 p(z_i) \tag{11.3.7}$$

是表示直方图偏斜度的量，而四阶矩是表示相关平直度的量。五阶矩和更高阶矩不容易同直方图形状联系起来，但它们提供对纹理描述更进一步的量化。另外一些有用处的基于直方图的纹理量度包括"一致性"量度，如下式给出的：

$$U = \sum_{i=0}^{L-1} p^2(z_i) \tag{11.3.8}$$

和平均熵值量度，读者可以回顾有关的基本信息理论，或第 8 章的讨论。这里将其定义为：

$$e = -\sum_{i=0}^{L-1} p(z_i)\log_2 p(z_i) \tag{11.3.9}$$

因为 p 在区间[0,1]内有值并且这些值的和为 1，量度 U 对所有灰度级都相等(极大一致性)的图像有最大值并且从那里开始递减。熵是可变性的量度并且对于一个不变的图像其值为 0。

例 11.6　基于直方图的纹理量度

表 11.2 总结了前述对于图 11.22 中高亮显示的三类纹理值的量度。均值正好告诉我们每个区域的平均灰度级并且它只在作为亮度的大致概念时才是有用的，并且不是真实的纹理。标准差可以提供更为丰富的信息；数字显示第一块纹理在灰度级上比另两块纹理有更小的可变性(它较平坦)。使用这个量度清晰地显示出粗糙的纹理。正如我们所希望的，相同的注释同样适用于 R，因为它和标准差所衡量的事物本质上是相同的。三阶矩一般对确定直方图的对称性和是否向左(负值)或向右(正值)歪斜很有用。

表 11.2　图 11.22 中子图的纹理量度

纹理	均值	标准差	R(归一化的)	三阶矩	一致性	熵
平滑	82.64	11.79	0.002	−0.105	0.026	5.434
粗糙	143.56	74.63	0.079	−0.151	0.005	7.783
规则性	99.72	33.73	0.017	0.750	0.013	6.674

这个量使我们对图像的灰度级相对于均值是偏向暗的一边还是偏向亮的一边，有了一个大略的认识。至于纹理这一术语，只有在量度之间的变化很大对，三阶矩所给出的信息才有用处。考虑一下一致性量度，我们也可以断定第一个子图更为平滑一些(比其他的图像一致性更好)并且多数的随机抽样(最低一致性)与粗糙的纹理相一致，这并不让人觉得意

外。最后,熵值变化与一致性量度是反向的,因此从中也可以得到同一致性量度相同的结论。第一个子图在灰度级上变化最小,而粗糙图像变化最大。规则纹理的灰度级变化在上述两种量度中处于两个极端之间。

仅使用直方图计算纹理的方法,由于直方图不具有像素之间相对位置的信息而受到限制。使纹理分析过程能得到这类信息的一个方法是,不仅考虑强度分布,还要考虑具有相同强度值或近似强度值的像素位置。

令 P 为位置算子并且令 $\mathbf{A}$ 为一个 $k \times k$ 阶矩阵。这个矩阵的元素 a_{ij} 是,出现(其位置由 P 指定)的灰度级为 z_i 的点相对于灰度级为 z_j 的点的倍数,$1 \leqslant i, j \leqslant k$。例如,考虑一幅具有3个灰度级的图像($z_1 = 0, z_2 = 1, z_3 = 2$)如下所示:

$$\begin{matrix} 0 & 0 & 0 & 1 & 2 \\ 1 & 1 & 0 & 1 & 1 \\ 2 & 2 & 1 & 0 & 0 \\ 1 & 1 & 0 & 2 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{matrix}$$

将位置算子 P 定义为"在右下方的一个像素",生成下列的 3×3 阶矩阵 $\mathbf{A}$:

$$\mathbf{A} = \begin{bmatrix} 4 & 2 & 1 \\ 2 & 3 & 2 \\ 0 & 2 & 0 \end{bmatrix}$$

其中,比如,a_{11}(左上角)是灰度为 $z_1 = 0$ 的点出现在具有相同灰度的像素右下方的次数,a_{13}(右上角)是灰度为 $z_1 = 0$ 的点出现在灰度为 $z_3 = 2$ 的点右下位置的次数。$\mathbf{A}$ 的大小由输入图像可分辨的灰度级数目决定。因此,这一节中讨论的概念在应用时一般需要将亮度重新在一个小的灰度级范围内进行量化,以便使矩阵易于处理。

令 n 为图像中满足 P 的点对总数目(前述例子中 $n = 16$,是矩阵 $\mathbf{A}$ 中所有值的和)。如果一个矩阵 $\mathbf{C}$ 是通过用 n 除 $\mathbf{A}$ 中的每一个元素得到的,则 c_{ij} 为概率估计,这个概率是一对满足 P 的、具有值(z_i, z_j)的点的联合概率。矩阵 $\mathbf{C}$ 称为灰度级共生矩阵。因为 $\mathbf{C}$ 依赖于 P,所以要找到给出的纹理模式需要选择一个合适的位置算子。例如,用于前述例子的算子对于沿 $-45°$ 分布的恒等强度区域很灵敏(注意 $\mathbf{A}$ 中的最大值为 $a_{11} = 4$,部分是由于一条强度为0的点组成的条纹为 $-45°$ 走向)。更一般的问题是,为了分类计算 $\mathbf{C}$ 的区域纹理而分析给定的 $\mathbf{C}$ 矩阵。为了实现这个目的,有用处的一些描绘子如下所示:

1. 最大概率:

$$\max_{i,j}(c_{ij})$$

2. 元素差异的 k 阶矩:

$$\sum_i \sum_j (i-j)^k c_{ij}$$

3. 逆元素差异的 k 阶矩:

$$\sum_i \sum_j c_{ij}/(i-j)^k, i \neq j$$

4. 一致性：

$$\sum_i \sum_j c_{ij}^2$$

5. 熵：

$$-\sum_i \sum_j c_{ij} \log_2 c_{ij}$$

基本思想是想通过这些描绘子描述 **C** 的"内容"特性。例如，第一条性质表示对 P 的最强响应。第二个描绘子当 **C** 中的大值接近主对角线时具有相对低的值，因为 $(i-j)$ 的差在此处较小。第三个描绘子有相反的效果。第四个描绘子当 c_{ij} 都相等时有最大值。如前面注意到的，第五个描绘子是一个随机性的度量，当 **C** 的所有元素有最大的随机性时，此描绘子具有最大值。

使用这些描绘子的一个方法是，"教授"一个系统有代表性的描绘子对不同的纹理集进行赋值。然后，未知区域有什么样的纹理就由它的描绘子是否与系统存储的纹理相匹配决定。将在第 12 章中更详细地讨论匹配。

结构性方法

如在本节开始时提到的，纹理描述的第二个主要范畴是基于结构的概念。假设我们有形如 $S \to aS$ 的规则，这种形式的规则表明字符 S 可以被重写为 aS（例如，三次应用此规则可生成字串 $aaaS$）。如果 a 表示一个圆[图 11.23(a)]，并且赋予形如 $aaa\cdots$ 的串以"向右排布的圆"的意义，则规则 $S \to aS$ 可以生成如图 11.23(b)所示的纹理模式。

假设下一步给这个方案增加一些新的规则：$S \to bA$，$A \to cA$，$A \to c$，$A \to bS$，$S \to a$，这里 b 的存在表示"向下排布的圆"，c 的存在表示"向左排布的圆"。现在可以生成一个形如 $aaabccbaa$ 的串。这个串对应一个圆的 3×3 阶矩阵。更大的纹理模式，如图 11.23(c)所示，用相同的方式可以很容易地生成（然而注意，这些规则也可能生成非矩形的结构）。

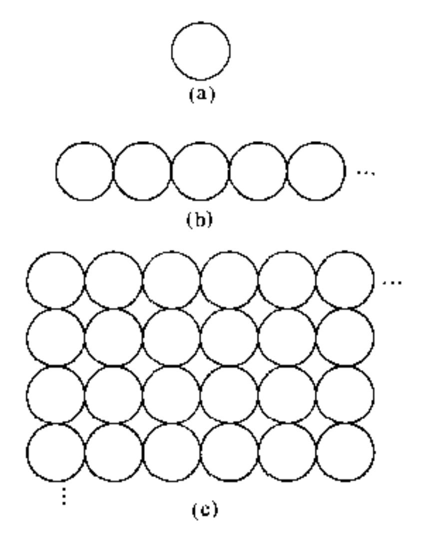


图 11.23　(a)纹理图元，(b)由规则 $S \to aS$ 生成的模式，(c)由 $S \to aS$ 和其他规则生成的二维纹理模式

上述讨论的基本思想是:一个简单的"纹理图元"可以借助一些限制图元可能的排列数规则组成更为复杂的纹理模式。这些概念是关系描绘子的核心,将在11.5节中详细讨论。

频谱方法

如5.4节中提到的,傅里叶频谱完全适合于描述图像中循环的方向性或循环的二维模式。这些整体的纹理模式,尽管作为频谱中高能脉冲的集中区域是易于区别的,但一般来讲,这些纹理模式由于处理技术的局部性,使用空间检测方法是十分困难的。

这里,考虑对纹理描述有用的傅里叶频谱的3个特征:(1)频谱中突起的尖峰给出了纹理模式的主要方向;(2)在频率平面中尖峰的位置给出了模式的基本空间周期;(3)通过过滤除去所有周期性的部分,而留下非周期性的图像元素,然后,这些留下的元素可以通过统计技术进行描述。回想一下,频谱对于原点是对称的,所以只需考虑半面频率平面即可。因此,为了进行分析,每个周期性模式都只与频谱中的一个尖峰相联系,而不是两个。

刚才提到的频谱特征的检测和解释通常使用函数 $S(r,\theta)$ 的极坐标表达比较简单。这里 S 是频谱函数,r 和 θ 是坐标系中的变量。对于每个方向 θ,$S(r,\theta)$ 可以看做一维函数 $S_\theta(r)$。类似地,对每个频率 r,$S_r(\theta)$ 也是一个一维函数。对固定的 θ 值分析 $S_\theta(r)$,可得到沿着自原点的辐射方向上的频谱所表现的特性(比如存在的尖峰)。反之,分析固定 r 值的 $S_r(\theta)$,可得到沿着以原点为圆心的圆形上的特性。

一种更具有整体性的描述通过对下列函数进行积分(对于离散变量为求和)得到:

$$S(r) = \sum_{\theta=0}^{\pi} S_\theta(r) \tag{11.3.10}$$

和

$$S(\theta) = \sum_{r=1}^{R_0} S_r(\theta) \tag{11.3.11}$$

这里 R_0 是以原点为圆心的圆半径。

式(11.3.10)和式(11.3.11)的结果为每对坐标 (r,θ) 组成一对值 $[S(r),S(\theta)]$。通过变换这些坐标,可以生成两个一维函数 $S(r)$ 和 $S(\theta)$,从而对研究的整幅图像或所考虑的区域纹理构成一种频谱-能量描述。再有,这些描绘子本身可以计算以便定量地描述它们自身的状态。典型的用于这种目的的描绘子是最高值的位置、均值、振幅和轴向偏差两者的方差,还有函数的均值和最高值之间的距离。

例11.7 频谱纹理

图11.24说明了式(11.3.10)和式(11.3.11)对于整体纹理描述的用处。图11.24(a)显示了一幅带有周期性纹理的图像,图11.24(b)显示了此图像的频谱。图11.24(c)和图11.24(d)分别显示了 $S(r)$ 和 $S(\theta)$ 的曲线。$S(r)$ 的曲线具有典型的结构,在原点附近有高能含量,及对于更高频率的渐进较低的值。$S(\theta)$ 的曲线显示出以45°为间隔的突出尖峰,明显与图像中纹理内容的周期性相对应。

作为 $S(\theta)$ 的曲线如何用于区别两个纹理模式的说明,图11.24(e)显示了另一幅在水平和垂直方向上具有较为突出的纹理模式的图像。图11.24(f)显示了这一图像频谱 $S(\theta)$ 的

曲线。正如我们希望的，此曲线显示了以 90°为间隔的尖峰。通过分析它们对应的 $S(\theta)$ 波形可以很容易地区分两个纹理模式。

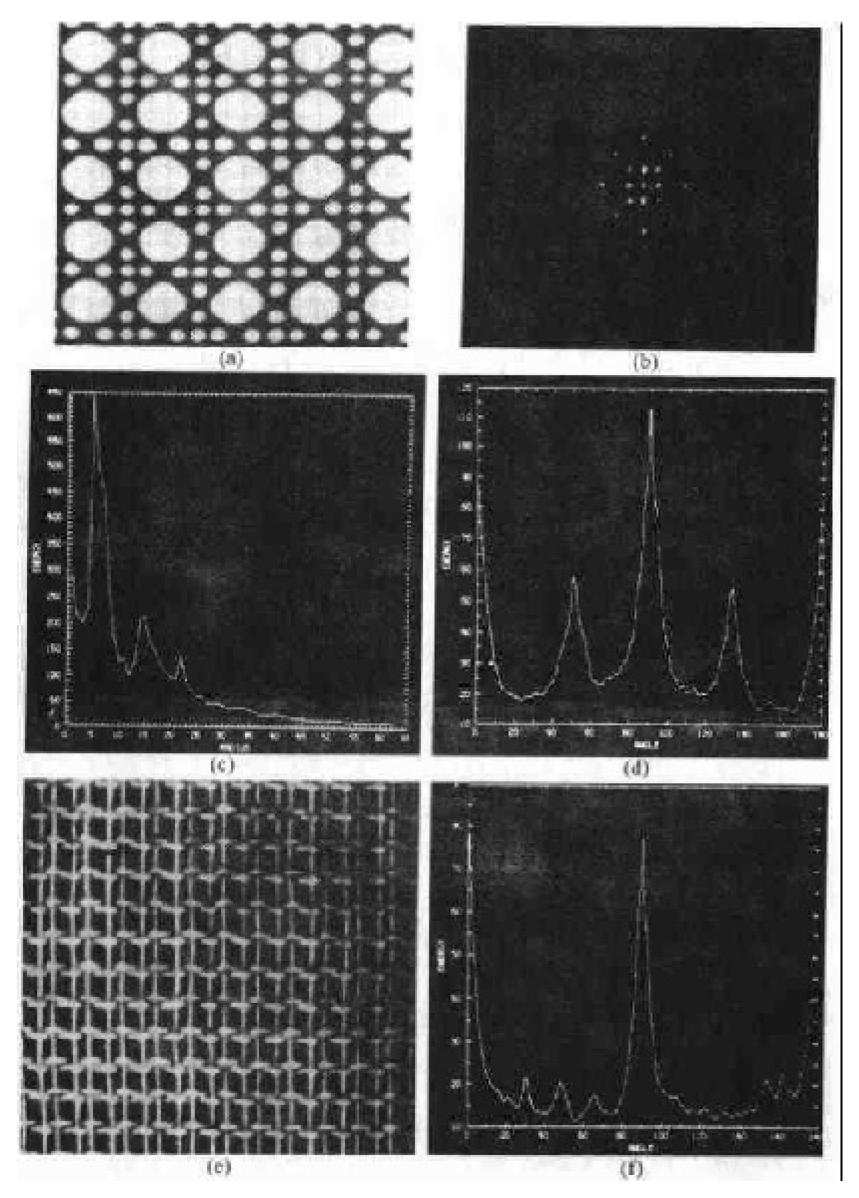


图 11.24　(a)显示了周期纹理的图像，(b)频谱，(c)$S(r)$的曲线，(d)$S(\theta)$的曲线，(e)具有不同类型的周期性纹理的另一幅图像，(f)$S(\theta)$的曲线(由田纳西大学的Dragana Brzakovic博士提供)

11.3.4　二维函数的矩

对于二维连续函数 $f(x,y)$，$(p+q)$阶矩定义为：

$$m_{pq}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x^p y^q f(x,y)\mathrm{d}x\mathrm{d}y \tag{11.3.12}$$

$p,q=0,1,2,\cdots$。由单值性定理(Papoulis[1991])说明：如果 $f(x,y)$是分段连续的并且仅在 xy 平面内有限的部分具有非零值，则存在各阶矩，并且矩的序列(m_{pq})由 $f(x,y)$惟一决定。相

反,(m_{pq})也惟一地决定 $f(x,y)$。

中心矩定义为:

$$\mu_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - x)^p(y - \bar{y}^q f(x,y)\mathrm{d}x\mathrm{d}y \tag{11.3.13}$$

这里,

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad 和 \quad \bar{y} = \frac{m_{01}}{m_{00}}$$

如果 $f(x,y)$是数字图像,则式(11.3.13)变成:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p(y - \bar{y})^q f(x,y) \tag{11.3.14}$$

总之,

$$\begin{aligned}
\mu_{00} &= \sum_x \sum_y (x - \bar{x})^0(y - \bar{y})^0 f(x,y) \\
&= \sum_x \sum_y f(x,y) \\
&= m_{00} \\
\mu_{10} &= \sum_x \sum_y (x - \bar{x})^1(y - y)^0 f(x,y) \\
&= m_{10} - \frac{m_{10}}{m_{00}}(m_{00}) \\
&= 0 \\
\mu_{01} &= \sum_x \sum_y (x - \bar{x})^0(y - \bar{y})^1 f(x,y) \\
&= m_{01} - \frac{m_{01}}{m_{00}}(m_{00}) \\
&= 0 \\
\mu_{11} &= \sum_x \sum_y (x - \bar{x})^1(y - \bar{y})^1 f(x,y) \\
&= m_{11} - \frac{m_{10}m_{01}}{m_{00}} \\
&= m_{11} - \bar{x}m_{01} = m_{11} - \bar{y}m_{10} \\
\mu_{20} &= \sum_x \sum_y (x - \bar{x})^2(y - \bar{y})^0 f(x,y) \\
&= m_{20} - \frac{2m_{10}^2}{m_{00}} + \frac{m_{10}^2}{m_{00}} \\
&= m_{20} - \frac{m_{10}^2}{m_{00}} \\
&= m_{20} - \bar{x}m_{10} \\
\mu_{02} &= \sum_x \sum_y (x - \bar{x})^0(y - \bar{y})^2 f(x,y)
\end{aligned}$$

$$= m_{02} - \frac{m_{01}^2}{m_{00}}$$
$$= m_{02} - \bar{y}m_{01}$$
$$\mu_{21} = \sum_x \sum_y (x - \bar{x})^2 (y - \bar{y})^1 f(x, y)$$
$$= m_{21} - 2\bar{x}m_{11} - \bar{y}m_{20} + 2\bar{x}^2 m_{01}$$
$$\mu_{12} = \sum_x \sum_y (x - \bar{x})^1 (y - \bar{y})^2 f(x, y)$$
$$= m_{12} - 2\bar{y}m_{11} - \bar{x}m_{02} + 2\bar{y}^2 m_{10}$$
$$\mu_{30} = \sum_x \sum_y (x - x)^3 (y - \bar{y})^0 f(x, y)$$
$$= m_{30} - 3\bar{x}m_{20} + 2\bar{x}^2 m_{10}$$
$$\mu_{03} = \sum_x \sum_y (x - \bar{x})^0 (y - \bar{y})^3 f(x, y)$$
$$= m_{03} - 3\bar{y}m_{02} + 2\bar{y}^2 m_{01}$$

三阶中心矩为：

$$\mu_{00} = m_{00} \qquad \mu_{02} = m_{02} - \bar{y}m_{01}$$
$$\mu_{10} = 0 \qquad \mu_{30} = m_{30} - 3\bar{x}m_{20} + 2\bar{x}^2 m_{10}$$
$$\mu_{01} = 0 \qquad \mu_{03} = m_{03} - 3\bar{y}m_{02} + 2\bar{y}^2 m_{01}$$
$$\mu_{11} = m_{11} - \bar{y}m_{10} \qquad \mu_{21} = m_{21} - 2\bar{x}m_{11} - \bar{y}m_{20} + 2\bar{x}^2 m_{01}$$
$$\mu_{20} = m_{20} - \bar{x}m_{10} \qquad \mu_{12} = m_{12} - 2\bar{y}m_{11} - \bar{x}m_{02} + 2\bar{y}^2 m_{10}$$

归一化中心矩，由 η_{pq} 表示，定义为：

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \tag{11.3.15}$$

这里，

$$\gamma = \frac{p + q}{2} + 1 \tag{11.3.16}$$

$p + q = 2, 3, \cdots$

一组 7 个不变矩来自于二阶和三阶矩①。

$$\phi_1 = \eta_{20} + \eta_{02} \tag{11.3.17}$$
$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \tag{11.3.18}$$
$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{11.3.19}$$
$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{11.3.20}$$
$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \tag{11.3.21}$$

① 这些结果的推导涉及的概念超出本节讨论的范畴。Bell[1965]著的书和 Hu[1962]写的论文中包含了这些概念的详细讨论。矩的不变式可以推广到 n 维空间(Mamistvalov[1998])。

$$[3(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2]$$

$$\phi_6=(\eta_{20}-\eta_{02})[(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2] + 4\eta_{11}(\eta_{30}+\eta_{12})(\eta_{21}+\eta_{03}) \tag{11.3.22}$$

$$\phi_7=(3\eta_{21}-\eta_{03})(\eta_{30}+\eta_{12})[(\eta_{30}+\eta_{12})^2 - 3(\eta_{21}+\eta_{03})^2]+(3\eta_{12}-\eta_{30})(\eta_{21}+\eta_{03}) [3(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2] \tag{11.3.23}$$

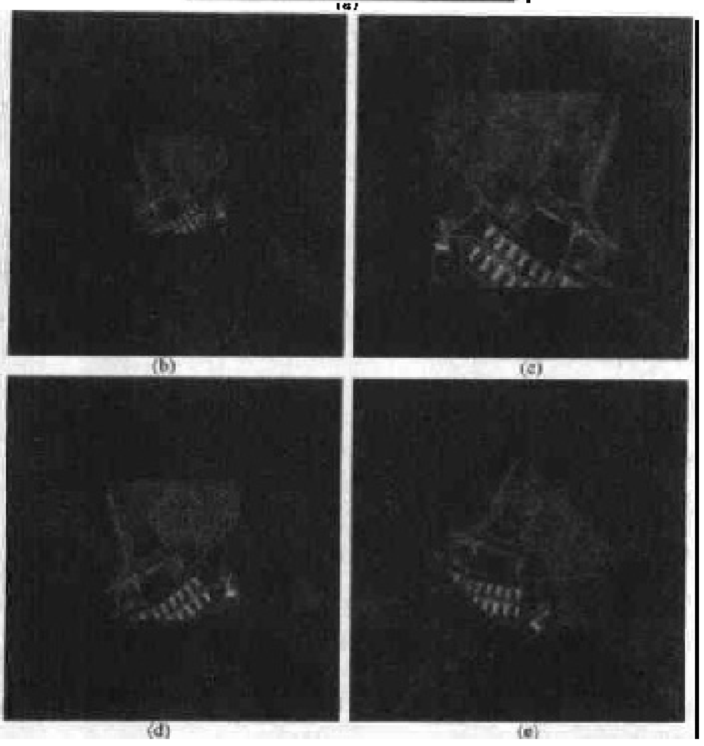
这些矩组对于平移、旋转和比例缩放是不变的。

例 11.8　二维不变矩

图 11.25(a)中显示的图像变为图 11.25(b)中的一半大小，图 11.25(c)中的是镜像图，图 11.25(d)中是旋转了 2°和 45°的图像。从式(11.3.17)到式(11.3.23)给出的 7 个不变矩都用于对这些图像进行计算，对其结果取对数是为了减小动态范围。如表 11.3 所示，图 11.25(b)到图 11.25(e)所得结果和对原图计算得到的不变矩有较好的一致性。产生误差的主要原因可归结为数据的数字化本质，特别是对于旋转的图像更是如此。



(a)



(b)　(c)　(d)　(e)

图 11.25　用于证明矩不变性质的图像(见表 11.3)

表 11.3　图 11.25(a)~(e)中图像的矩不变量

不变量(Log)	原值	一半尺寸	镜像	旋转 2°	旋转 45°
ϕ_1	6.249	6.226	6.919	6.253	6.318
ϕ_2	17.180	16.954	19.955	17.270	16.803
ϕ_3	22.655	23.531	26.689	22.836	19.724
ϕ_4	22.919	24.236	26.901	23.130	20.437
ϕ_5	45.749	48.349	53.724	46.136	40.525
ϕ_6	31.830	32.916	37.134	32.068	29.315
ϕ_7	45.589	48.343	53.590	46.017	40.470

11.4　运用主分量进行描绘

本节中提到的资料适用于边界和区域。另外,这些资料也可用做图像组描述的基础,这些图像已做过空间配准,但是这个集合对应的像素值是不同的(比如彩色 RGB 图像的 3 个分量图像)。假设得到了这样一幅彩色图像的 3 个分量图像。这 3 幅图像可以通过将每组 3 个对应的像素表示成一个向量而看做一个单元。例如,令 x_1, x_2 和 x_3 分别为这 3 幅图像的第 1 个像素的值。这 3 个元素可以用三维列向量 $\mathbf{x}$ 的形式表示,这里有:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

这个向量表示在所有 3 幅图像中的一个普通像素。如果图像大小为 $M \times N$,将所有像素用这种方式表示后,将有总数为 $K = MN$ 的三维向量。如果有 n 个配准了的图像,向量将是 n 维的:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{11.4.1}$$

纵观这一节,我们假定所有的向量都是列向量(即 $n \times 1$ 阶矩阵)。可以把它们写成一行文本,即 $\mathbf{x} = (x_1, x_2, \cdots, x_n)^T$ 的形式,这里"T"表示转置。

可以把向量当做随机量,就像我们构造灰度级直方图时做的那样。仅有的不同点是,现在谈论的是随机向量的均值向量和方差矩阵,而不是随机变量的均值和方差。总体的均值向量定义为:

$$\mathbf{m}_x = E\{\mathbf{x}\} \tag{11.4.2}$$

这里 $E\{\cdot\}$是变量的期望值,脚标表示 $\mathbf{m}$ 与向量 $\mathbf{x}$ 的总体相联系。向量或矩阵的期望值是通过取每个元素的期望值得到的。

总体向量的方差矩阵定义为:

$$\mathbf{C_x} = E\{(\mathbf{x} - \mathbf{m_x})(\mathbf{x} - \mathbf{m_x})^T\} \tag{11.4.3}$$

因为 $\mathbf{x}$ 是 n 维的，$\mathbf{C_x}$ 和$(\mathbf{x} - \mathbf{m_x})(\mathbf{x} - \mathbf{m_x})^T$ 是$n \times n$ 阶矩阵。$\mathbf{C_x}$ 的元素 c_{ii}是总体向量中 $\mathbf{x}$ 向量的第 i 个分量x_i 的方差，而且 $\mathbf{C_x}$ 的元素 c_{ij}是这些向量元素 x_i 和x_j 的协方差。[①] 矩阵 $\mathbf{C_x}$ 是实对称的。如果元素 x_i 和x_j 无关，它们的协方差为零，且有 $c_{ij} = c_{ji} = 0$。注意，当 $n = 1$ 时，所有这些定义都降为常见的一维对应量。

对于从随机总体中取样的 K 向量，均值向量可以通过使用常见的求平均值的表达式由样本来近似得到。

$$\mathbf{m_x} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{x}_k \tag{11.4.4}$$

同样，通过扩展乘积$(\mathbf{x} - \mathbf{m_x})(\mathbf{x} - \mathbf{m_x})^T$ 并使用式(11.4.2)和式(11.4.4)，会发现协方差矩阵可以以如下方式用样本近似得到：

$$\mathbf{C_x} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{x}_k\mathbf{x}_k^T - \mathbf{m_x}\mathbf{m_x}^T \tag{11.4.5}$$

例 11.9 均值向量和协方差矩阵的计算

为了说明式(11.4.4)和式(11.4.5)的机制，考虑 4 个向量 $\mathbf{x}_1 = (0,0,0)^T$，$\mathbf{x}_2 = (1,0,0)^T$，$\mathbf{x}_3 = (1,1,0)^T$和 $\mathbf{x}_4 = (1,0,1)^T$，这里使用转置以便列向量可以像前面注释的一样方便地写成横向文本。应用式(11.4.4)得到下列均值向量：

$$\mathbf{m_x} = \frac{1}{4}\begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$$

同样，使用式(11.4.5)得到下列协方差矩阵：

$$\mathbf{C_x} = \frac{1}{16}\begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix}$$

所有沿着主对角线的元素都相等，这表示总体向量中向量的 3 个分量有同样的方差。元素 x_1 和 x_2，x_1 和 x_3 必定相关；元素 x_2 和 x_3 必定不相关。

因为 $\mathbf{C_x}$ 是实对称的，找到一组 n 个标准正交特征向量总是可能的(见 Noble 和 Daniel [1988]的著作)。令 $\mathbf{e}_i$ 和λ_i，$i = 1,2,\cdots,n$，为特征向量和对应的$\mathbf{C_x}$特征值，以降序排布(为了方便)使 $\lambda_j \geqslant \lambda_{j+1}$，$j = 1,2,\cdots,n-1$。[②] 令 $\mathbf{A}$ 为一个由 $\mathbf{C_x}$ 的特征向量组成其行元素的矩阵，并进行排序，使 $\mathbf{A}$ 的第一行为对应最大特征值的特征向量，而最后一行为对应最小特征值的特征向量。

假设将 $\mathbf{A}$ 作为将 $\mathbf{x}$ 的向量映射到用 $\mathbf{y}$ 代表的向量的变换矩阵，如下所示：

---

① 回顾具有均值为 m 的随机变量 x 的方差定义为 $E\{(x-m)^2\}$，两个随机变量 x_i 和 x_j 的协方差定义为 $E\{(x_i - m_i)(x_j - m_j)\}$。如果变量不相关，它们的协方差为 0。

② 根据定义，$n \times n$ 矩阵 $\mathbf{C}$ 的特征向量和特征值满足 $\mathbf{C}\mathbf{e}_i = \lambda_i\mathbf{e}_i$，$i = 1,2,\cdots,n$。

$$\mathbf{y} = \mathbf{A}(\mathbf{x} - \mathbf{m}_x) \tag{11.4.6}$$

则这个表达式称为霍特林(Hotelling)变换，这个变换就像将在下面简要说明的那样，具有某些有趣的和有用的性质。

不难说明从这一变换得到的 $\mathbf{y}$ 向量的均值是 0，即：

$$\mathbf{m}_y = E\{\mathbf{y}\} = 0 \tag{11.4.7}$$

这一结论是依据基本矩阵理论：关于 $\mathbf{y}$ 向量的协方差矩阵是由 $\mathbf{A}$ 和 $\mathbf{C}_x$ 用下列表达式得到的。

$$\mathbf{C}_y = \mathbf{A}\mathbf{C}_x\mathbf{A}^T \tag{11.4.8}$$

再有，因为 $\mathbf{A}$ 的构成方式，$\mathbf{C}_y$ 是一个对角矩阵，其主对角线上的元素是 $\mathbf{C}_x$ 的特征值；即：

$$\mathbf{C}_y = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix} \tag{11.4.9}$$

这个协方差矩阵的非对角线上的元素为 0，所以向量组 $\mathbf{y}$ 的元素是不相关的。记住，λ_j 是 $\mathbf{C}_x$ 的特征值，并且沿着对角矩阵主对角线的元素是 $\mathbf{C}_x$ 的特征值(见 Noble 和 Daniel[1988]的著述)。因此，$\mathbf{C}_x$ 和 $\mathbf{C}_y$ 有相同的特征值。实际上，它们的特征向量也是相同的。

霍特林变换的另外一个重要性质是处理由 $\mathbf{y}$ 重构 $\mathbf{x}$ 的问题。因为 $\mathbf{A}$ 的各行是正交向量，具有 $\mathbf{A}^{-1} = \mathbf{A}^T$ 的性质，任何向量 $\mathbf{x}$ 能够通过相应的 $\mathbf{y}$ 使用下列表达式恢复回来：

$$\mathbf{x} = \mathbf{A}^T\mathbf{y} + \mathbf{m}_x \tag{11.4.10}$$

然而，假设不使用 $\mathbf{C}_x$ 的特征向量，而从 k 个特征向量对应的 k 个最大特征值得到一个 $k \times n$ 转换矩阵，从而构成矩阵 $\mathbf{A}_k$。向量组 $\mathbf{y}$ 则成为 k 维的，并且也不再需要式(11.4.10)的重构方法(这与我们在 11.2.3 节中应用的使用少量傅里叶系数描述一个边界的过程有些相似)。

使用 $\mathbf{A}_k$ 重构向量的方法如下：

$$\hat{\mathbf{x}} = \mathbf{A}_k^T\mathbf{y} + \mathbf{m}_x \tag{11.4.11}$$

$\mathbf{x}$ 和 $\hat{\mathbf{x}}$ 之间的均方误差可以由下列表达式给出：

$$\begin{aligned} e_{ms} &= \sum_{j=1}^{n}\lambda_j - \sum_{j=1}^{k}\lambda_j \\ &= \sum_{j=k+1}^{n}\lambda_j \end{aligned} \tag{11.4.12}$$

式(11.4.12)的第一行表明，如果 $k = n$(即，如果所有的特征向量都用于变换)，则误差为零。因为 λ_j 单调减少，式 (11.4.12)也说明误差可以通过选择 k 个具有最大特征值的特征向量而降至最小。因此，从可以将向量 $\mathbf{x}$ 和它的近似值 $\hat{\mathbf{x}}$ 之间的均方误差降至最小这方面来说，霍特林变换是最佳的。由于这种使用特征向量对应的最大特征值的思想，霍特林变换也称为主分量变换。

例 11.10 使用主分量描述图像

图 11.26 显示了由一个六波段多光谱扫描器生成的 6 幅图像,所使用的波段列在表 11.4 中。像本节开始时提到的那样,观察图 11.27 所示的图像,这些图像中每组对应的像素可以构成一个六维向量 $\mathbf{x}=(x_1,x_2,\cdots,x_6)^T$。在这个特定的应用事例中,图像分辨率为 384×239,因此总体包含有 91 776 个向量用于计算均值向量和协方差矩阵。表 11.5 显示了 $\mathbf{C}_x$ 的特征值。注意,前两个特征值的优势。

使用式(11.4.6)生成一组对应于 x 向量的 y 向量的变换。根据这些向量,6 个主要的分量图像被组合在一起(图像由简单地应用图 11.27 中反向的向量重构)。图 11.28 显示了结果。分量 1 代表由进行过向量转换的所有 y_1 的分量形成的图像,其他 5 幅图像也是如此。回想一下基本矩阵理论,例如,通过用列向量$(\mathbf{x}-\mathbf{m}_x)^T$ 计算 $\mathbf{A}$ 的第一行的内积(点积)得到 y_1。

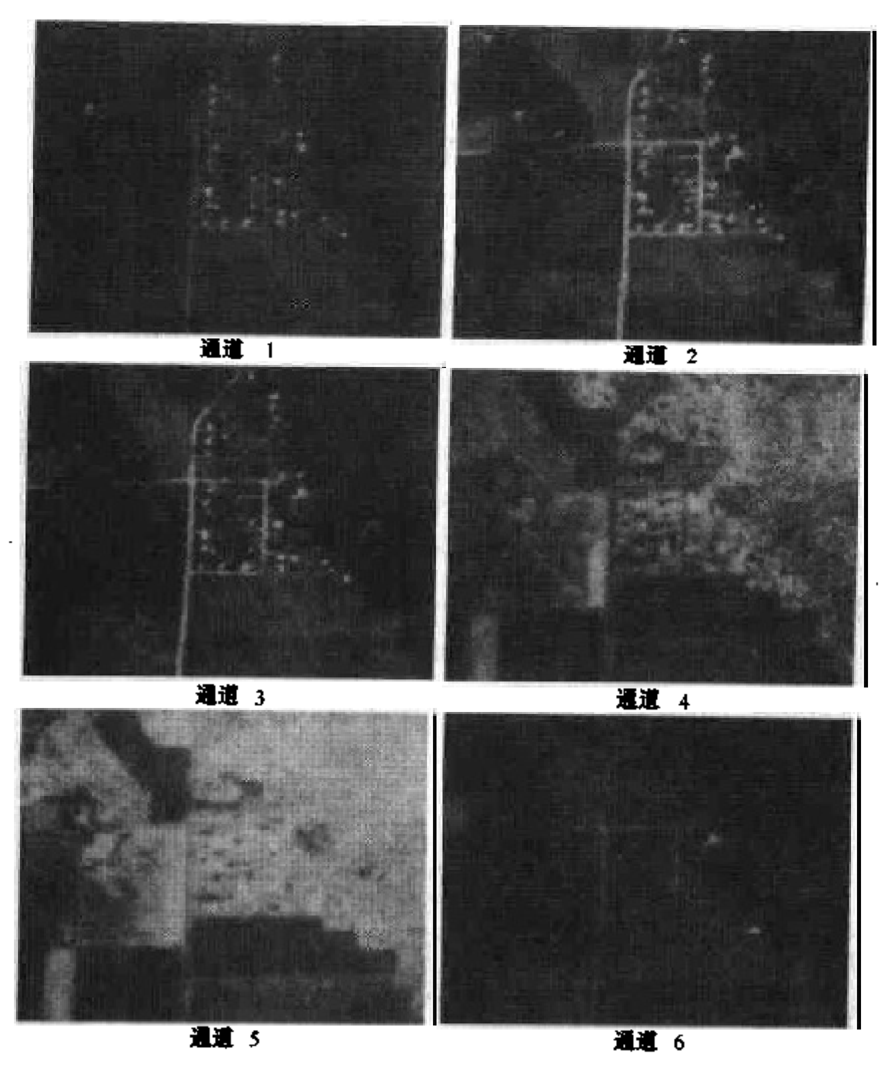


图 11.26 来自机载扫描器的 6 幅光谱图像(由 Purdue 大学遥感应用实验室提供)

表 11.4　通道编号和对应的波长

通　道	波段(μm)
1	0.40～0.44
2	0.62～0.66
3	0.66～0.72
4	0.80～1.00
5	1.00～1.40
6	2.00～2.60

A 的第一行是特征向量,它对应于总体协方差矩阵的最大特征值,并且这个特征值给出了第一幅经过变换的图像的灰度级方差。因此,基于表 11.5 中显示的数字,这幅图具有最高的对比度。这个结果在图 11.28 中是相当明显的。因为前两幅图像的方差值占了总体方差值的 94%,所以,另 4 幅主分量图像对比度低的事实是意料之中的事。因此,如果仅将变换后的前两幅图连同 $\mathbf{m}_x$ 和 **A** 的前两行进行存储,而不是为了以后处理而存储所有的 6 幅图像,那么为了可靠一些,在存储之后生成一幅与原来的 6 幅图像近似的图像是必需的。这种进行数据压缩的能力尽管以当今的标准衡量并未给人以深刻印象,但却是霍特林变换产生的一项有用的副产品。从图像描述的方面来讲,这种方法的意义在于可以用两幅图像加上均值向量和变换矩阵的前两行对 6 幅图像的内容进行描述。如果我们要论述的是区域而不是整幅图像,则同样的办法是可以应用的。

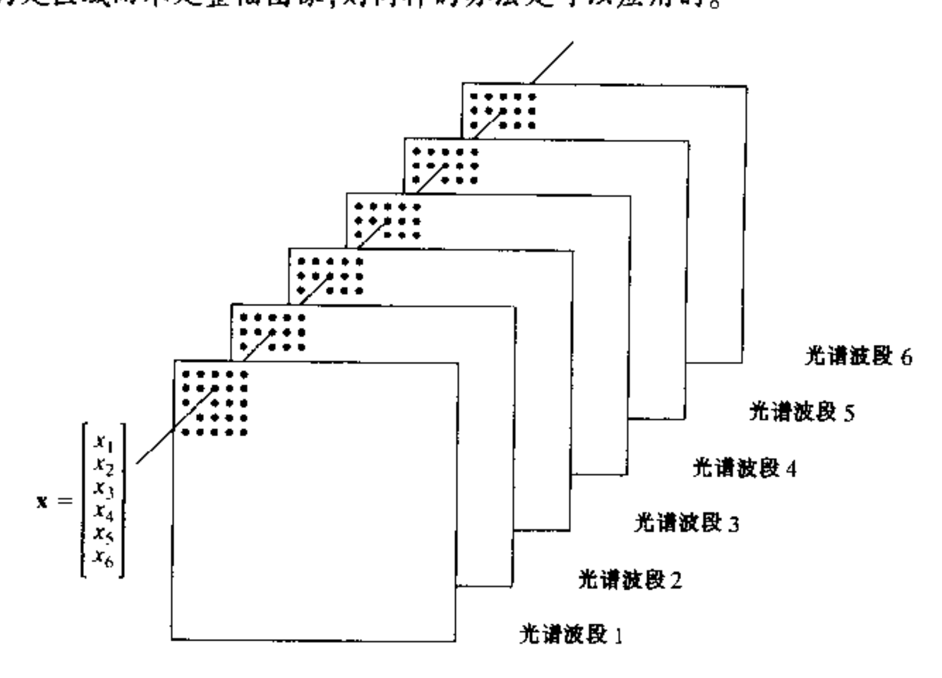


图 11.27　根据 6 幅图像对应的像素点构成的一个向量

表 11.5　从图 11.26 中的图像得到的协方差矩阵特征值

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
3210	931.4	118.5	83.88	64.00	13.40

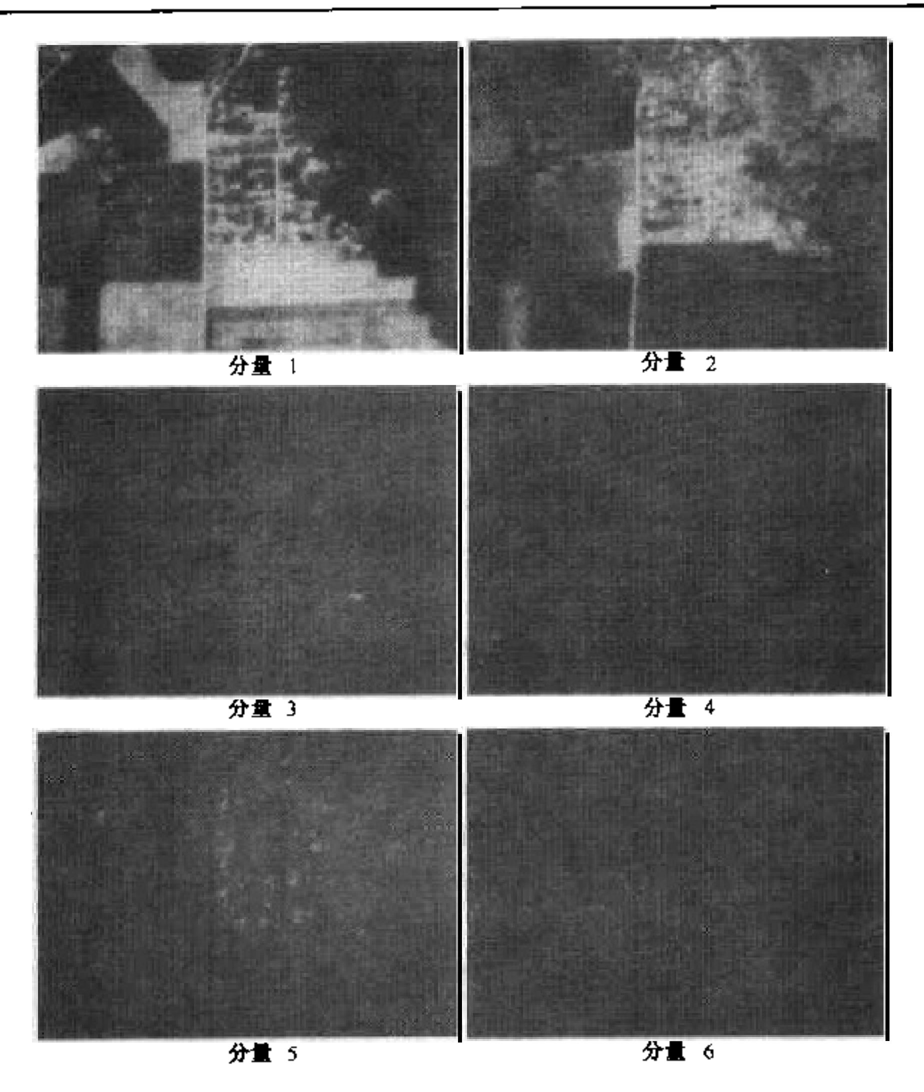


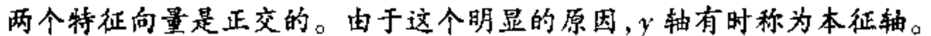
图 11.28 根据图 11.26 的数据计算得到的 6 幅主分量图像(由 Purdue 大学遥感应用实验室提供)

例 11.11 在单幅图像中使用主分量对边界和区域进行描述

在前面的讨论中,说明了如何对一组图像或区域使用主分量变换。在这个例子中,用图示说明如何使用主分量描述单幅图像中的边界和区域。方法是根据边界或区域的坐标生成二维向量。考虑图 11.29(a)中显示的对象。如果希望描述这个区域,就根据对象中像素点坐标生成向量。如果希望描述边界,就只使用边界上的点坐标。

结果向量作为二维随机向量的总体处理。换句话说,对象中的每个像素点作为一个二维向量 $\mathbf{x}=(a,b)^T$ 去处理,这里 a 和 b 是像素关于 x_1 轴和 x_2 轴的坐标值。这些向量用于计算总体(对象)的均值向量和协方差矩阵。由于只在二维中计算,所以问题比以前简单了许多。

使用式(11.4.6)的实际结果是需要设置一个新的坐标系统,这个坐标系统以总体(均值向量的坐标)的质心为原点,以 $\mathbf{C}_x$ 的特征向量所指方向为轴的方向,如图 11.29(b)所示。这个坐标系统清晰地显示出式(11.4.6)所进行的变换是一种旋转变换,这种变换使用特征向量将数据排列起来,如图 11.29(c)所示。实际上,这种排列正好是数据去相关的机理。再有,由于特征值沿着 $\mathbf{C}_y$ 的主对角线排布,λ_i 是沿着特征向量 $\mathbf{e}_i$ 的分量 y_i 的方差。这


两个特征向量是正交的。由于这个明显的原因，y 轴有时称为本征轴。

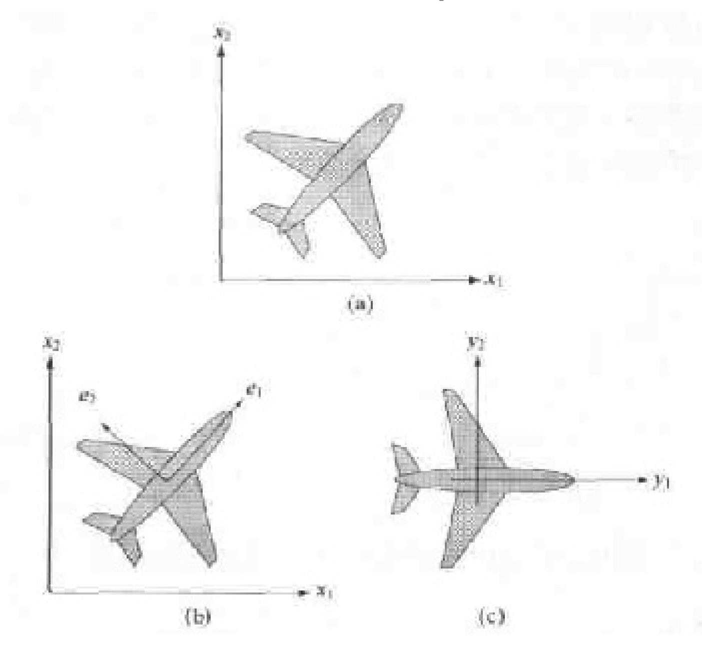


图 11.29　(a)一个对象，(b)特征向量，(c)使用式(11.4.6)使对象旋转。实际效果是将对象沿着自身的本征轴对准

使用主特征向量排列二维对象的概念在图像描述中起着十分重要的作用。如前面提到的，图像描述对于大小变化、平移和旋转变化本应是尽可能独立的。使用对象的主轴矫正的能力为消除旋转变化的影响提供了一种可靠手段。特征值是沿着本征轴的方差，并可用于尺寸的归一化。平移带来的影响可以通过将对象的均值设定为中心来解决，如式(11.4.6)所说明的。请记住，本节中推导的描述方法对区域和边界是同样适用的。

11.5　关系描绘

在 11.3.3 节中为描述纹理而介绍了重写规则的概念。本节中，在关系描绘子的内容中推广这一概念。这些概念对边界或区域同样适用，并且它们的主要目的是以重写规则的形式在边界和区域中获取基本的重复模式。

考虑如图 11.30(a)中所示的简单阶梯结构。假设这个结构是从一幅图像中分割出来的，并且希望用某种形式化的方法对它进行描述。通过定义如图所示的两个图元元素 a 和 b，可以以图 11.30(b)所示的形式对图 11.30(a)进行编码。这个编码结构的最明显性质就是元素 a 和 b 的重复。所以，一种简单的描述方法是用公式表达一种包含这些图元元素的递归关系。一种可能是使用重写规则：

(a) $S \rightarrow aA$

(b) $A \rightarrow bS$

(c) $A \rightarrow b$

这里 S 和 A 是变元，元素 a 和 b 是对应于刚才定义的图元的常元。规则 1 表明 S(称为起始符)可以被图元 a 和变元 A 代替。依次地，这个变元 A 可以被 b 和 S 或只被 b 代替。用 bS 代

替 A,然后回到第一条规则,并且重复这一过程。用 b 代替 A 结束这一过程,因为在表达式中不再存在变元。图 11.31 说明了一些由这些规则推导的例子,这个结构下面的数字表示规则 1,2,3 应用的次序。a 和 b 的关系被保存了,因为这些规则使一个 b 总是跟在一个 a 后面。值得注意的是,这 3 个简单的重写规则可以用于生成(或描述)无限多"相同的"结构。如将在第 12章中说明的,这种方法的优点有坚实的理论基础。

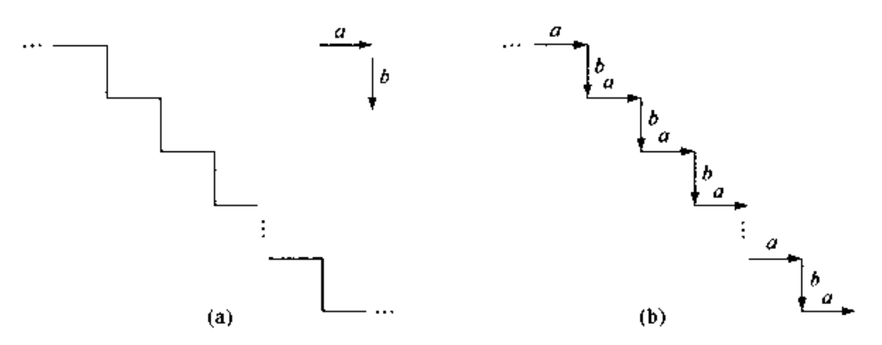


图 11.30 (a)一个简单的阶梯结构,(b)编码结构

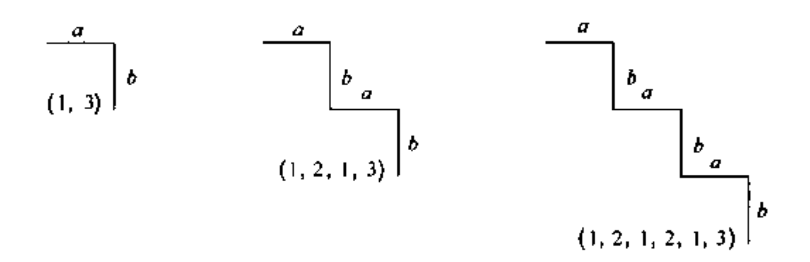


图 11.31 用规则 $S \to aA$, $A \to bS$ 和 $A \to b$ 推导的例子

因为串是一维结构,所以将串应用于图像描述需要建立一种适当的方法,将二维的位置关系转变为一维形式。大多数使用串对图像进行描述的应用都是基于从有用对象中抽取连接线段的思想。一种方法是沿着对象的轮廓线用指定了方向和长度的线段进行编码。图 11.32 说明了这一过程。

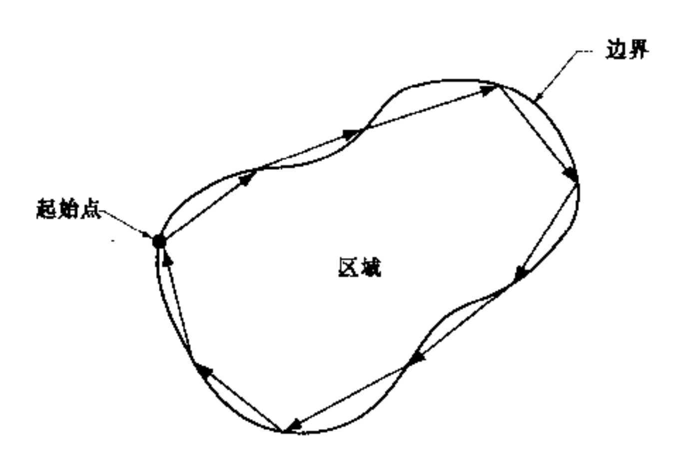


图 11.32 用定向线段对一个区域边界进行编码

另一种更为普遍一些的方法是,通过定向线段描述一幅图像(如小的单色区域)的部分区域,这种定向线段除了首尾连接关系外可以通过其他方法连接起来。图 11.33(a)说明了这一

方法,图 11.33(b)显示了一些典型操作,这些操作可以用于定义抽象图元。图 11.33(c)显示了一组由 4 个方向定义的线段组成的特定图元。图 11.33(d)显示了一个特定图形生成的步骤,这里(~ d)表示与图元 d 相同但方向相反的图元。注意每个复合结构只有一个头和一个尾。其有意义的结果就在于最后一个串,这个串描述了完整的结构。

串描述最适合应用于以首尾连接或其他连接方式表达的图元连通性情况。有时,从纹理或其他描绘子看来相似的区域却是不连接的,这就需要有描述这种情况的技术。用于处理这种情况的最有效方法之一是使用树描绘子。

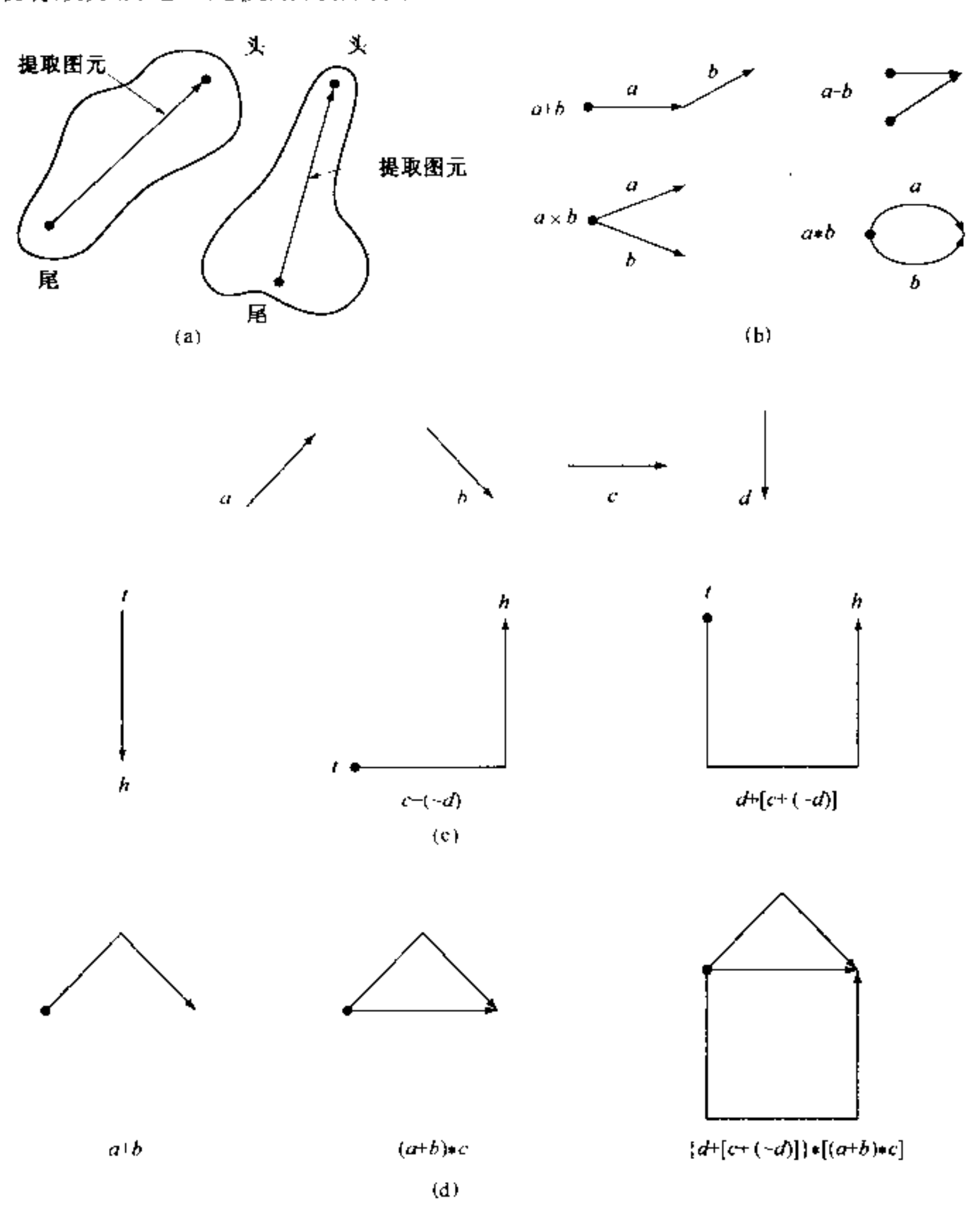


图 11.33　(a)一个提取的图元,(b)图元间的操作,(c)一组特定的图元,(d)生成一个结构的步骤

一棵树 T 是一个或多个节点的有限集合：

(a)仅有惟一一个根节点 \$，并且

(b)余下节点被分成彼此不连接的集合 $T_1,\cdots,T_m$(依次均为一棵树，称为 T 的子树)。树的末梢节点是在树的底端从左到右依次排列的节点集合(叶子)。例如，图 11.34 显示的树有根 \$和末梢 xy。

一般来说，树中的两类信息是重要的：(1)关于节点的信息，它是以描述节点的一组字来存储的，(2)一个节点与它的相邻节点关系的信息，它是以指向那些相邻节点的指针集合来存储的。

正如图像描述中那样，第一类信息表示一幅图的结构(例如，区域或边界线段)，而第二类信息定义了一个结构和其他结构间的物理关系。例如，图 11.35(a)可以用树通过使用"在……之内"的关系进行描绘。因此，如果树的根用 \$表示，图 11.35(a)显示出第一级关系，即 a 和 c 包含于 \$，这种关系产生两个由根节点发出的分支，如图 11.35(b)所示。下一级关系包括 b 在 a 内，d 和 e 在 c 内。最后，由 f 在 e 内完成整棵树。

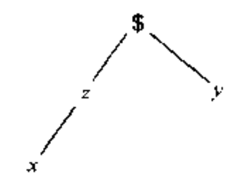


图 11.34 带有根节点 \$和末梢节点 xy 的简单树

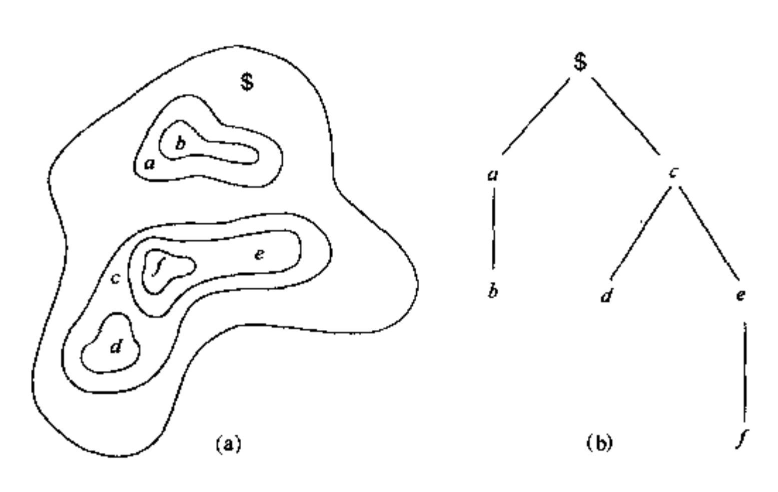


图 11.35 (a)一个简单的复合区域，(b)通过使用关系"在……之内"得到的树描绘

小结

从图像中分割出来的对象或区域的表示方法和描绘方式是大多数图像自动处理过程的前期步骤。例如，这些描绘构成了将在下一章中展开讨论的对象识别方法的输入。正如这一章描绘技术所涵盖范围指出的那样，从各种方法中挑选哪一种方法是由所面对的问题决定的。

目的就是选择能够"抓住"对象或对象类之间本质差异的描绘子,而对于各种因素的变化,比如位置、尺寸和方向,应尽可能地保留这种差异的独立性。

参考资料

11.1.1 节中讨论的链码表示是由 Freeman[1961,1974]首次提出的。近年来有关使用链码的研究可以见 Bribiesca[1999]的著述,他将链码推广到了三维空间(Bribiesca[2000])。边界的多边形描绘(见 11.1.2 节)由于它有用而简单很值得考虑。

十年前 Bengtsoon 和 Eklundh[1991]及 Sato[1992]的论文对这一领域中典型的研究工作进行了论述。Zhu 和 Chirlian[1995]写的论文论述了一种有意义的沿着曲线变形点检测的方法。这种方法也可以参见 Hu 和 Yan[1997]的著述。最近这一领域的研究工作集中于不变量的多边形拟合(参见 Voss 和 Suesse[1997]的著述)和多边形近似算法(参见 Rosin[1997]的著述)的效果评估、一般实现(参见 Huang 和 Sun[1999]的著述)及计算速度(参见 Davis[1999]的著述)。

符号(见 11.1.3 节)的讨论可参考 Ballard 和 Brown[1982]及 Gupta 和 Strinath[1988]的著述。关于寻找凸壳和凸形缺陷(见 11.1.4 节)的基本公式表示可参见 Preparata 和 shamos[1985]的著述。也可参考 Liu-Yu 和 Antipolis[1993]的论文。Katzir 等人[1994]讨论了部分封闭曲线的检测。Zimmer 等人[1997]讨论了计算凸壳的一个改进算法,Latecki 和 Lakämper[1999]讨论了图形分解的凸性法则。

11.1.5 节中讨论的骨架化算法基于 Zhang 和 Suen[1984]的著述。另外一些关于这个算法的性质和实现的有用论述包括在 Lu 和 Wang[1986]论文中。Jang 和 Chin[1990]的论文在 11.1.5 节中的讨论和 9.5.5 节中介绍的关于细化的形态学概念之间提供了有机的联系。对于存在噪声的细化方法见 Shi 和 Wong[1994]以及 chen 和 Yu[1996]的著述。Shaked 和 Bruckstein[1998]讨论了一种修剪算法,对消除骨架中的痕迹非常有用。中轴变换的快速计算在 Sahni 和 Jenq[1992]以及 Ferreira 和 Ubéda[1999]的著述中进行了讨论。由 Loncaric[1998]写的关于 11.1 节中讨论的许多方法的综述文章也很有意义。

Freeman 和 Shapira[1975]给出了寻找一个闭合链码曲线的基本矩形算法(见 11.2.1 节)。11.2.2 节中的关于形状数目的讨论是基于 Bribiesca 和 Guzman[1980]以及 Bribiesca[1981]的工作。对于傅里叶描绘子(11.2.3 节)的其他读物见 Zahn 和 Roskies[1972]以及 Persoon 和 Fu[1077]的早期论文。也可以参见 Aguado 等人[1998]和 Sonka 等人[1999]的著述。Reddy 和 Chatterji[1996]讨论了使用 FFT 算法实现平移、旋转和比例缩放变换的不变性的重要方法。11.2.4 节中的材料基于基本的概率理论(例如,参见 Peebles[1993]和 Popoulis[1991]的著述)。

关于 11.3.2 节的其他读物,参见 Rosenfeld 和 Kak[1982]以及 Ballard 和 Brown[1982]的著述。关于纹理(11.3.3 节)的精辟介绍参见 Haralick 和 Shapiro[1992]的著述。早期的关于纹理的综述参见 Wechsler[1980]的著述。Murino 等人[1998]和 Garcia[1999]的论文以及 Shapiro 和 Stockman[2001]所做的论述是当前这一领域中有代表性的研究成果。

11.3.4 节中讨论的矩不变方法来自 Hu[1962]的著述。也可参见 Bell[1965]的著述。要了解矩不变量应用范围的概念,参见 Hall[1979]关于图像匹配的论述和 Cheung 与 Teoh[1999]关于使用矩描述对称性的论述。Mamistvalov[1998]的论述将矩的不变量方法推广到 n 维空间。

Hotelling[1933]是第一个推导并发表将离散变量转换成不相关系数这一方法的人。他把这种技术称为主分量方法。他的论文对这种方法有相当的见地,值得一读。霍特林变换被

Kramer 和 Mathews[1956]以及 Huang 和 Schultheiss[1963]再次发现。主分量方法在大量应用中依旧是用于图像描绘的基本工具,在 Swets 和 Weng[1996]以及 Duda, Heart, Stork[2001]的著述中举有这方面的例子。11.5 节中的参考材料来自 Gonzalez, Thomason[1978]和 Fu[1982]的著述,也可参见 Sonka 等人[1999]的著述。

习题

11.1★(a)说明链码起点的重定义方法以使得到的编号序列构成一个具有最小量级的整数。这种重定义使编码不依赖于边界上的初始起点。

(b)找出编码 11076765543322 的归一化起点。

11.2 (a)像 11.1.1 节解释的那样,说明链码的首差使其对于旋转变化实现归一化。

(b)计算编码 0101030303323232212111 的首差。

11.3★(a)说明 11.1.2 节中讨论的橡皮带多边形近似方法生成一个具有最小周长的多边形。

(b)说明如果每个单元对应一个边界上的像素,那么单元中最大可能的误差为$\sqrt{2}d$,这里 d 是相邻像素之间的最小可能水平或垂直距离(即,用于生成数字图像的取样网格的直线间距离)。

11.4★(a)如果将 11.1.2 节中讨论的聚合方法的误差门限值设为零,讨论对得到的多边形的影响。

(b)对分割方法有什么影响?

11.5★(a)使用 11.1.3 节讨论的正切角方法划分方形边界的图。

(b)用斜率密度函数重复上述划分。

假设这个方形的边分别与 x 轴和 y 轴形成一条直线,x 轴为基准线。以最靠近原点的点为起点。

11.6 为下列每条边界图形找到一个表达式,并划分图形。

★(a)一个等边三角形

(b)一个矩形

(c)一个椭圆形

11.7 画出下列图形的中轴

★(a)一个圆

★(b)一个方形

(c)一个矩形

(d)一个等边三角形

11.8 对下页每个显示的图形:

★(a)讨论在点 p 执行 11.1.5 节中介绍的骨骼化算法的第一步时所执行的操作。

(b)重复算法的第二步。假设在所有情况下 $p=1$。

1	1	0
1	p	0
1	1	0

0	0	0
1	p	0
0	0	0

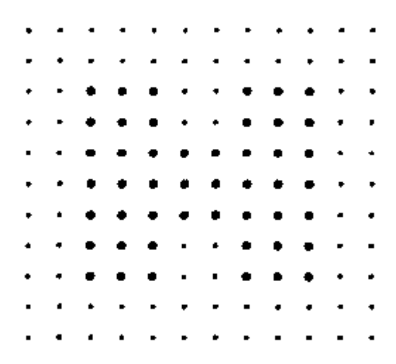
0	1	0
1	p	1
0	1	0

1	1	0
0	p	1
0	0	0

11.9 参考 11.1.5 节中介绍的骨骼化算法,在执行下列操作后图形是什么样子?

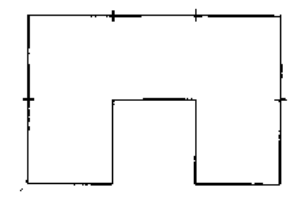
★(a)执行一遍算法的第一步。

(b)执行一遍第二步(在第一步的结果上执行,不是对原图像)。



11.10★(a)对所显示的图形,形状数的阶数是什么?

(b)写出形状数。



11.11 11.2.3 节中讨论的使用傅里叶描绘子的过程用复数表示轮廓坐标,写出这些复数的 DFT,并仅保留一些 DFT 的分量作为边界图形的描绘子。逆 DFT 是对原始轮廓线的近似。哪一类轮廓图形具有由实数构成的 DFT,并且如何设置图 11.13 中的轴系统才能得到这些实数?

★**11.12** 给出区分图 11.5 中显示的图形所需的最小数目的统计矩描绘子。

11.13 给出两个具有相同均值和三阶统计矩描绘子但二阶统计矩不同的边界图形。

★**11.14** 提出一组能区分字符 0,1,8,9 和 X 形状的描绘子(提示:将拓扑描绘子和凸壳一起使用)。

11.15 考虑一个由黑白方块交错组成的跳棋棋盘图形,每个方块大小为 $m \times m$。给出可以生成一个对角的共生矩阵的位置算子。

11.16 画出 5×5 大小的图像的灰度共生矩阵,图像由交错的 1 和 0 的跳棋棋盘图案组成。

★(a)位置算子 P 定义为"右边的一个像素"和

(b)"右边的两个像素"。

假设左上角的像素具有0值。

11.17 证明式(11.4.7)、式(11.4.8)和式(11.4.9)的正确性。

★11.18 在例11.10中提到,用两幅与最大特征值相联系的主分量图像重构6幅原图的近似图像是较为可靠的做法。这样做均方误差会受到什么影响?将答案表示成一个最大可能误差的百分比。

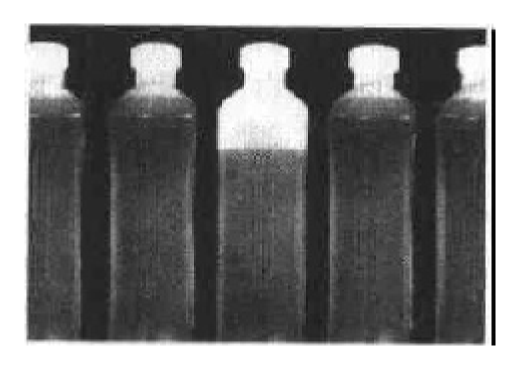
11.19 对一组大小为64×64的图像,假设式(11.4.9)给出的协方差矩阵证明是单位矩阵。则原图和仅用原特征向量的一半由式(11.4.11)重构的图像间的均方误差如何表示?

★11.20 如果11.2.1节定义的边界主轴和其本征轴相等,希望有什么条件?

11.21 给出一个由黑白相间的方块组成的棋盘图形的空间关系及对应的树形表达。假设左上角元素为黑色且树根对应此元素。构造的树从每一个节点辐射的分支不要多于两个。

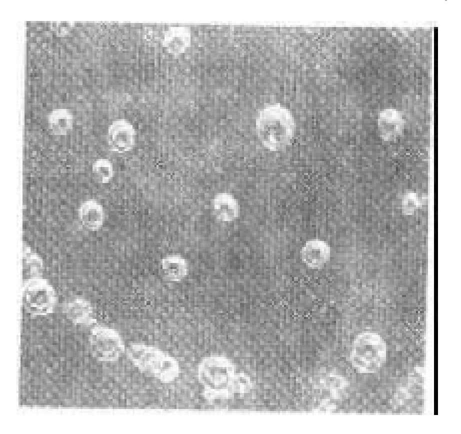
★11.22 要求设计一个检测某种固态晶体圆片内部缺陷的图像处理系统。晶片使用X射线成像系统检测,产生分辨率为512×512的8位图像。在没有缺陷的地方图像显示为平均灰度级是100及方差是400的"均匀区"。有缺陷的地方呈现如斑点状的区域。这个区域有大约70%的像素偏离50这一灰度级或均值小于100。如果有缺陷的区域所占面积超过20×20个像素则认为此晶片有缺陷。对此提出一个基于纹理分析的系统。

11.23 一个用瓶子装各种工业化学品的装瓶公司听说你成功解决了成像问题,便雇佣你设计一种检测瓶子未装满的方法。当瓶子在传送带上运动并通过自动装填机和封盖机时如下图所示。当液体平面低于瓶颈底部和瓶子肩部的中间点时,认为瓶子未装满。瓶子的横断面上的倾斜部分及侧面定义为瓶子的肩部。瓶子在不断移动,但公司有一个图像系统,装备了有效捕捉停止运动图像的前端闪光照明设备。所以可以得到非常接近于这里显示的样例的图像。基于以上得到的资料,提出一个检测未完全装满的瓶子的解决方案。清楚地表述所做的所有很可能对你提出的解决方案产生影响的假设。



11.24 听说你成功地解决了装瓶问题,你和一家公司签订合同,这家公司希望在某个过程中自动进行气泡计数以便进行质量监控。他们已经解决了成像问题并可以得到分辨率为700×700像素的8位数据图像,如下页所示。每幅图像代表7 cm^2 的面积。公司希望用每幅图做两件事:(1)确定图像中气泡的面积占图像总面积的

比率,和(2)计算可分辨的气泡数目。基于以上提供的资料,提出一个解决方案。在你的解决方案中,务必阐述方案中可检测的最小气泡的物理尺寸。清楚地说明所做的所有很可能对你提出的解决方案产生影响的假设。



第 12 章 对象识别

这里通过对对象识别技术的介绍总结一下数字图像处理所涵盖的范围。正如在 1.1 节注意到的,我们已经定义了包括识别单个图像区域在内的数字图像处理过程所涵盖的范围。在本章中我们称单个图像区域为对象或模式。

在这一章中阐述的模式识别方法分为两个主要方面:理论判别和结构判别。第一类方法处理的是使用定量描绘子描述的各种模式。定量描绘子有:长度、面积和纹理,等等。第二类方法处理的是大部分由定性描绘子描述的各种模式。比如,11.5 节讨论的关系描绘子。

识别的中心问题是从样本模式"学习"这一概念。在接下来的材料,我们将对研究理论判别和结构判决两类方法的"学习"技术进行深入讨论并给出图示加以说明。

12.1 模式和模式类

模式即描绘子的组合,例如那些在第 11 章中讨论过的符号。特征在有关模式识别的著作中经常被用于表示一个描绘子。模式类是一个拥有某些共同性质的模式族。模式类用 $\omega_1,\omega_2,\cdots,\omega_W$ 表示,在这里 W 是模式类的数量。由机器完成的模式识别包括对不同的模式分配各自所属类的技术——这种技术是自动的并且尽可能地减少人的介入。

实践中三种常用的模式组合是矢量(用于定量描述)、串和树(用于结构描述)。模式矢量用粗体小写字母表示,比如:**x**,**y** 和 **z**,并采取下列形式:

$$\mathbf{x}=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{12.1.1}$$

这里,每个分量 x_i 代表第 i 个描绘子,n 是与模式有关的符号总数。模式矢量用列向量表示(即 $n\times 1$ 阶矩阵)。因此模式矢量可以用式(12.1.1)中所示的形式或用公式 $\mathbf{x}=(x_1,x_2,\cdots,x_n)^T$ 的等价形式表示,这里 T 表示矩阵转置。读者可以在 11.4 节了解这种表示法。

模式矢量 **x** 中的元素性质取决于描述物理模式自身所采用的方法。让我们用一个简单的例子来说明并给出度量分类的历史意义。在一篇经典的论文中,费希尔(Fisher [1936])发表了后来称为判别式分析的新技术(在 12.2 节讨论),它通过测量花瓣的宽度和长度(图 12.1)来识别三种不同类型的鸢尾属(Iris)植物的花(Iris Setosa, Iris Virginica 和 Iris Versicolor)。

在已有的术语中,每一种花都用两个度量描述,这就生成了一个二维模式矢量的形式。

$$\mathbf{x}=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{12.1.2}$$

x_1 和 x_2 分别代表花瓣的长度和宽度。本例中的 3 个模式类 $\omega_1,\omega_2,\omega_3$ 分别表示 Setosa,

Virginica和 Versicolor 三种花。

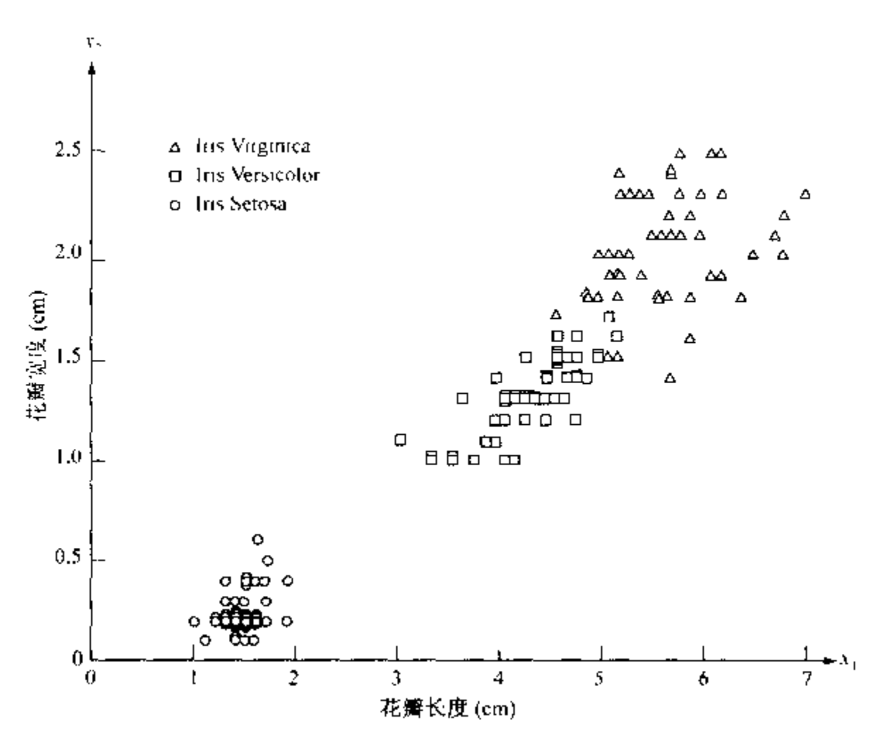


图 12.1 用两个度量描述三类鸢尾属植物的花

由于花瓣在宽度和长度上不同,描述这些花的模式矢量也会不同,这种不同不仅存在于不同种类的花瓣之间,在同一种类的花瓣间也存在。图 12.1 显示了每一类鸢尾属植物花瓣样本的长度和宽度的度量。在选定了一系列度量后(本例中选择两组数据),模式矢量的元素便成为每一个物理样本的完整描述。因此,例子中的每一朵花就由二维欧氏空间中的一个点代表。我们也注意到,本例中的花瓣长度和宽度较好地将 Iris Setosa 类同其他两类区分开来,但对 Virginica 和 Versicolor 类则分辨效果不佳。这种结果说明了经典特征选择方法的问题,说明类型可分离程度的高低很大程度上取决于应用的描绘子的选择。对此我们将在 12.2 节和 12.3 节中做更多的阐述。

图 12.2 显示了另一个生成模式矢量的例子。在这个例子中,关注的是不同类型的噪声波形。其中一个样本示于图 12.2(a)。如果我们选择用噪声信号代表每一个对象(见 11.1.3 节),会得到如图 12.2(b)中显示的一维信号形式。假设选择样本信号的幅度值来描述每一个信号,即以指定的间隔值 θ 对信号进行取样,表示为 $\theta_1,\theta_2,\cdots,\theta_n$。然后可以通过令 $x_1=r(\theta_1),x_2=r(\theta_2),\cdots,x_n=r(\theta_n)$构成模式矢量。这些矢量成为 n 维欧氏空间中的点,并且模式类被画成 n 维空间中的"云团"。

如果不直接使用信号的幅度,可以计算一个给定信号(见 11.2.4 节)的前 n 个统计矩,并使用这些描绘子作为每一个模式矢量的分量。实际上,现在很明显,可以有很多其他方法生成模式矢量。将在这一章中介绍其中的一些方法。目前,应记住的关键概念是,一个模式矢量的元素是根据什么量的描绘子进行选择的,这对于基于模式矢量方法进行对象识别的最终效果有很大影响。

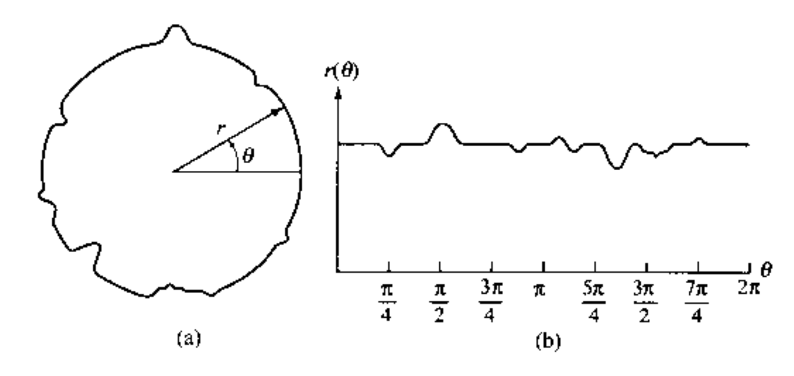


图 12.2　一个噪声对象和它对应的信号

刚才描述的这种产生模式矢量的技术可生成出定量信息描述其特性的模式类。在某些应用中,模式的特性很适于用结构关系进行描述。例如,指纹识别是基于称为细节的指纹特征的相互关系。综合指纹的相对大小和位置,这些特征是描述指纹纹路属性的主要分量,如指纹的端点、分支、合并以及不连续段。这类识别问题通常用结构性方法会得到很好的解决,因为它们的特征不仅同数量有关,而且各个特征间的空间关系也决定着它们的类别归属。此问题将在 11.5 节中介绍。在这里我们重新简略地回顾了有关模式描绘子的内容。

图 12.3(a)显示了一个简单的阶梯状模式。这个模式可以从模式矢量的角度取样和表达,如同在图 12.2 中使用的方法。然而,用这种描述方法将使其基本的结构特征(即两个原始元素构成的重复出现的特征)无法表现出来。一种更有意义的描述方法是定义元素 a 和 b 并将模式定义成如图 12.3(b)中所示的字符串 $w = \cdots abababab \cdots$。这种特殊模式类别的结构是由要求的描述得到的,这种描述用一种首尾方式定义连接,并且只允许交换符号。这种结构构成适用于任何长度的阶梯,但不包括由原始 a 和 b 连接成的另一类字串生成的其他类型结构。

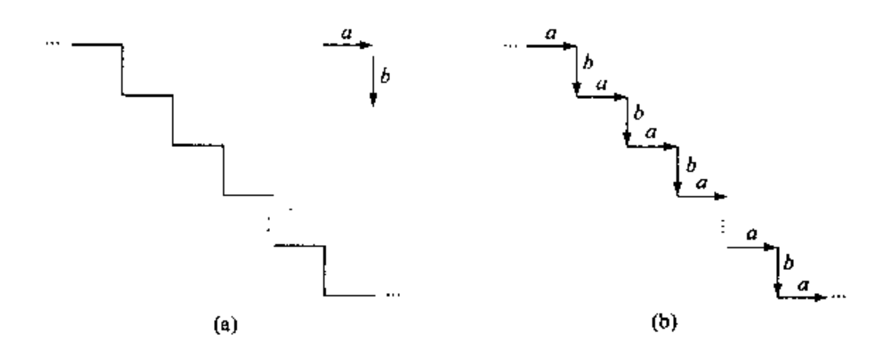


图 12.3　(a)阶梯结构,(b)基于元素 a 和 b 的结构编码生成串描述$\cdots ababab \cdots$

串的描述适于生成对象模式和其他实体模式,它们的结构是基于原始元素的较简单连接,通常和边界形状有关系。对许多应用来说更有效的一种方法是树形描述结构,如在 11.5 节中定义的。基本上,树形结构是一种主要的分层有序结构。例如,图 12.4 显示了一幅有稠密建筑物群的城市中心区及周围居民区的卫星照片。用 $定义整个图像区域。图 12.5中显示的树形表示法(从上到下)是通过应用结构关系中的"包含于"得到的。因此,树的根结点代表整幅图像。下一级的结点表示此图是由城市中心区和居民区构成的。居民区依次由住宅、高速公路和商业区构成。再下一层进一步描述了住宅和高速公路。可以一直继续这样的细分,直到到达在图像上解析不同区域的能力极限。

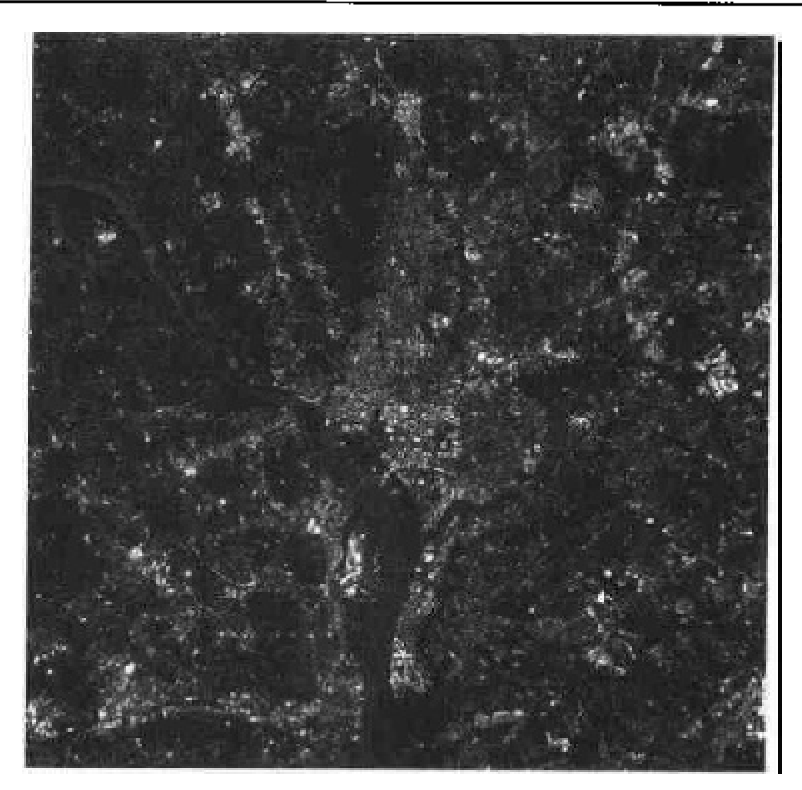


图 12.4 有大量建筑物的城市中心区(华盛顿特区)和周围的居民区的卫星图片(由 NASA 提供)

我们在接下来的章节中将深入探讨在以前的段落中所讨论技术描述的对象识别方法。

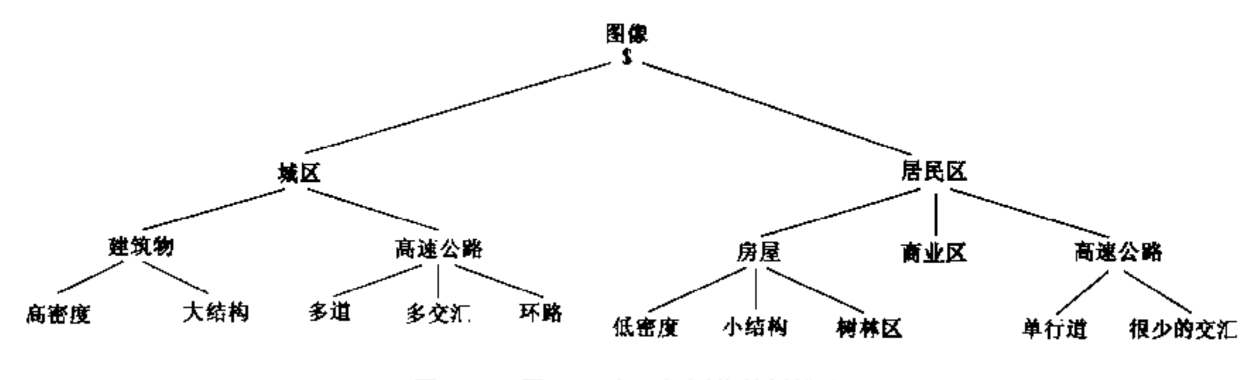


图 12.5 图 12.4 中一幅图像的树描述

12.2 基于决策理论方法的识别

使用决策理论方法的识别是基于使用决策(或判别)函数的识别。令 $\mathbf{x}=(x_1,x_2,\cdots,x_n)^T$ 代表一个 n 维模式矢量,如在 12.1 节中讨论的。对 W 个模式类 $\omega_1,\omega_2,\cdots,\omega_W$,决策理论模式识别的基本问题是依据属性寻找 W 个判别函数 $d_1(\mathbf{x}),d_2(\mathbf{x}),\cdots,d_W(\mathbf{x})$,如果模式 $\mathbf{x}$ 属于类 ω_i,则

$$d_i(\mathbf{x})>d_j(\mathbf{x})\quad j=1,2,\cdots,W;j\neq i \tag{12.2.1}$$

换句话说,一个未知的模式 $\mathbf{x}$ 被称为属于第 i 个模式类,只有当把 $\mathbf{x}$ 代入所有判别函数后能得到 $d_i(\mathbf{x})$ 的最大值时才成立。被解决的模式和模式类之间的联系可以是任意的。

从 ω_j 中分离 ω_i 类别的决策边界由 $\mathbf{x}$ 值判断，$\mathbf{x}$ 值由 $d_i(\mathbf{x}) = d_j(\mathbf{x})$ 或等价地由下式给出：

$$d_i(\mathbf{x}) - d_j(\mathbf{x}) = 0 \tag{12.2.2}$$

通常是通过单一函数 $d_{ij}(\mathbf{x}) = d_i(\mathbf{x}) - d_j(\mathbf{x}) = 0$ 来识别两个类之间的决策(判别)边界。因此，$d_{ij}(\mathbf{x}) > 0$时属于 ω_i 类的模式，而 $d_{ij}(\mathbf{x}) < 0$ 时属于 ω_j 类的模式。本节重点探讨各种各样的寻找判别函数的方法。这些函数要满足式(12.2.1)。

12.2.1 匹配

基于匹配的识别技术通过一种原型模式矢量表示每一个类。一个未知模式被赋予一个按预先已定义的度量与其最相近的类。最简单的方法是最小距离分类器，这种方法如其名称所暗示的一样，要计算(在欧几里得空间中的)未知量和每一个原型矢量间的距离。选择其中的最小距离来决策。还讨论过一种基于相关的方法，这种方法可以直接依据图像用公式表示并且相当直观。

最小距离分类器

假设我们把每一个模式类的原型定义为该类模式的平均矢量：

$$\mathbf{m}_j = \frac{1}{N_j}\sum_{x\in\omega_j}\mathbf{x}_j \quad j = 1,2,\cdots,W \tag{12.2.3}$$

这里 N_j 是类 ω_j 的模式矢量数目，在总和式中取全部这些矢量。如以前一样，W 是模式类的数目。决定一个未知模式矢量 $\mathbf{x}$ 的类别成员的方法是，如前边所述的那样，将它分配给与它最接近的原型类。使用欧几里得空间距离判断紧密性简化了计算距离测度的问题：

$$D_j(\mathbf{x}) = \|\mathbf{x} - \mathbf{m}_j\| \quad j = 1,2,\cdots,W \tag{12.2.4}$$

这里 $\|\mathbf{a}\| = (\mathbf{a}^T\mathbf{a})^{1/2}$ 是欧几里得范数。$D_i(\mathbf{x})$ 的值是最小距离时，把 $\mathbf{x}$ 划归给类 ω_i。也就是说，最小距离暗示此公式表示了最好匹配。不难看出(习题 12.2)选择最小距离等同于下面的函数评估：

$$d_j(\mathbf{x}) = \mathbf{x}^T\mathbf{m}_j - \frac{1}{2}\mathbf{m}_j^T\mathbf{m}_j \quad j = 1,2,\cdots,W \tag{12.2.5}$$

并且在 $d_i(\mathbf{x})$ 得出最大数值时将 $\mathbf{x}$ 划归给类 ω_i。这个公式同式 (12.2.1)定义的决策函数的概念一致。

在式(12.2.2)和式(12.2.5)中，类 ω_i 和 ω_j 之间的最小距离分类器的决策边界是：

$$\begin{aligned} d_{ij}(\mathbf{x}) &= d_i(\mathbf{x}) - d_j(\mathbf{x}) \\ &= \mathbf{x}^T(\mathbf{m}_i - \mathbf{m}_j) - \frac{1}{2}(\mathbf{m}_i - \mathbf{m}_j)^T(\mathbf{m}_i - \mathbf{m}_j) = 0 \end{aligned} \tag{12.2.6}$$

由式(12.2.6)给出的面是连接 $\mathbf{m}_i$ 和 $\mathbf{m}_j$ 线段的中垂线。$n = 2$ 时，中垂线是一条线；$n = 3$ 时，是平面；$n > 3$ 时，是一个超平面。

例 12.1　最小距离分类器的图解

图 12.6 显示了两个从图 12.1 鸢尾属植物例子中抽象出来的模式类。这两个模式类：Iris

Versicolor和Iris Setosa分别由ω_1和ω_2表示，样本平均矢量$\mathbf{m}_1=(4.3,1.3)^T$和$\mathbf{m}_2=(1.5,0.3)^T$。根据式(12.2.5)，决策函数是：

$$d_1(\mathbf{x})=\mathbf{x}^T\mathbf{m}_1-\frac{1}{2}\mathbf{m}_1^T\mathbf{m}_1$$
$$=4.3x_1+1.3x_2-10.1$$

和

$$d_2(\mathbf{x})=\mathbf{x}^T\mathbf{m}_2-\frac{1}{2}\mathbf{m}_2^T\mathbf{m}_2$$
$$=1.5x_1+0.3x_2-1.17$$

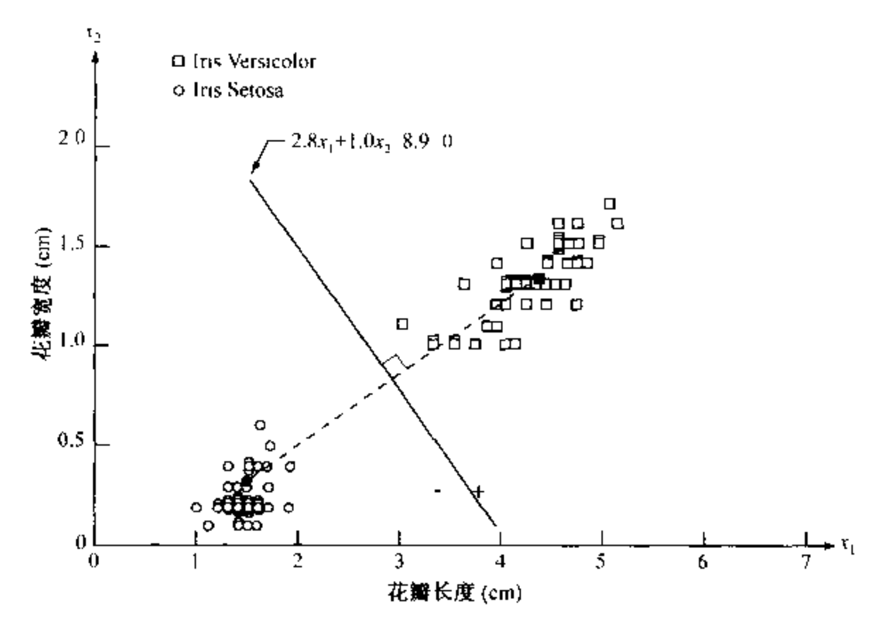


图 12.6 Iris Versicolor和Iris Setosa类的最小距离分类器的决策边界。黑点和方块是平均值

由式(12.2.6)得出边界方程是：

$$d_{12}(\mathbf{x})=d_1(\mathbf{x})-d_2(\mathbf{x})$$
$$=2.8x_1+1.0x_2-8.9=0$$

图12.6显示了具有这种边界的一幅图(注意轴线并未使用同一比例)。任何模式矢量代入类ω_1会得到$d_{12}(\mathbf{x})>0$。相反，任何模式矢量代入类ω_2会得到$d_{12}(\mathbf{x})<0$。换句话说，给出一个属于这两个类之一的未知模式，依据$d_{12}(\mathbf{x})$的符号将完全可以判断模式归属的类。

实际上，就每个模式类的均值而论，当均值间的间距与每一类的分开度或随机性相比很大时，最小距离分类器的分辨效果是很好的。在12.2.2节，我们可以看到，当每个类的均值分布在n维模式空间中呈现一种球形的“超级云团”状态时，最小距离分类器会得到最佳性能(从错误分类的平均损失的最小水平上来说)。

除系统的设计者控制输入性质之外，在实际中，大的均值分离和相对小的类散布同时出现是很少见的。一个很好的例子就是设计系统来读取有固定格式特征的字体。例如，大家所熟悉的美国Banker协会的E-13B字体的字符集。如图12.7所示，这个特殊的字符集由14个字

符组成,这些字符被特意设计在 9×7 个格子中以便于读取。这些字符通常用含有精细磁性材料的墨水印刷。在读取这些字符之前,墨水经过一个磁场,它加重每个字符以便简化检测。换句话说,分割问题已经用人工突出每个字符的关键特征解决了。

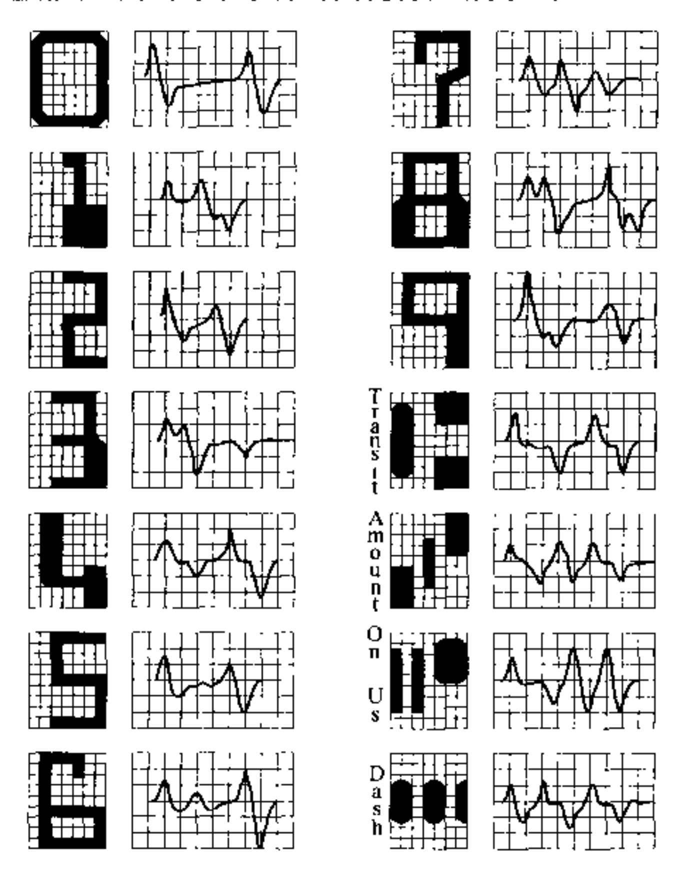


图 12.7　美国 Banker 协会的 E-13B 字体的字符集和对应波形

典型的字符扫描是在水平方向上用一个比字符长但更窄的单缝读取头进行的。当读取头扫过一个字符时,它产生一个一维的电信号,这个电信号随读取头下方的字符面积的增加或减少按比率变化。例如,考虑图 12.7 中数字 0 有关的波形。当读取头从左向右运动时,从读取头看去的字符区开始增大,从而产生一个正的变形(一个正的变化率)。当读取头开始远离 0 的左半部时,读取头下方的字符面积开始减少,产生一个负的变形。当读取头处于字符的中间区域时,字符面积近乎保持不变,产生一个 0 变形。当读取头进入字符的右半部分时,这种变形模式重复出现。字体的设计保证每个字符产生与其他字符不同的波形。同时也保证每种波形的波峰和零点大致在网格背景的垂直线上出现,这些波形同图 12.7 所显示的一样。E-13B 字体有一个特性,即仅仅在某些点采样就可以产生足够的恰当分类的特征信息。磁性墨水的使用有助于产生干净的波形从而将采样的散布减少到最小。

为这类应用设计一个最小距离分类器很简单。只需要存储每种波形的样本值,而后令每组样本用一个原型矢量 $\mathbf{m}_j(j=1,2,\cdots,14)$ 来表示。当对一个未知字符进行分类时,按刚才所述的方法去扫描它,将波形的网格样本表示成一个矢量 $\mathbf{x}$,选择式(12.2.5)中得到最大值的原型矢量类,这个类就是矢量 $\mathbf{x}$ 所属的类。由电阻排组成的模拟电路可以实现高速分类(见习题 12.4)。

依据相关的匹配

在 4.6.4 节中介绍了图像相关的基本概念。把这一概念作为在一幅大小为 $M \times N$ 的图像 $f(x,y)$ 中寻找匹配的大小为 $J \times K$ 的子图 $w(x,y)$ 的基础。这里,我们假定 $J \leqslant M$ 且 $K \leqslant N$。尽管相关方法可以用矢量形式表达(见习题 12.5),但直接对一幅图像或子图形式进行操作更为直观(也更为常见)。

在其最简单的表示形式中,$f(x,y)$ 和 $w(x,y)$ 间的相关是:

$$c(x,y) = \sum_s \sum_t f(s,t)w(x+s,y+t) \tag{12.2.7}$$

$x = 0,1,2,\cdots,M-1,y = 0,1,2,\cdots,N-1$,和式代表图像 w 和 f 相互重叠的地方。请注意将本式同式 (4.6.30)相比较,式 (4.6.30)暗示假设函数都是实的量并且忽略常数 MN。原因是将用一种能将常数抵消掉的归一化函数,并且式 (12.2.7)中给出的定义在实际使用中更为常见。并且在式 (12.2.7)中使用符号 s 和 t 避免同 m 和 n 相混淆,这些符号在本章中也用于其他目的。

图 12.8 用图示说明了处理步骤,这里,假设 f 原点在它的左上角,w 原点在它的中心。对于 $f(x,y)$ 中 x,y 的一个值,比如说(x_0,y_0),应用式(12.2.7)可得到量 c 的一个值。当 x 和 y 的值变化时,w 在图像区中移动,给出函数 $c(x,y)$。c 的最大值指出 w 能够最好地匹配 f 的位置。注意,在 x 和 y 的值接近 f 的边缘处时,c 值会失去准确性。在相关中误差量与 w 的大小成比例。这是我们熟悉的边界问题,在第 3 章已多次遇到过。

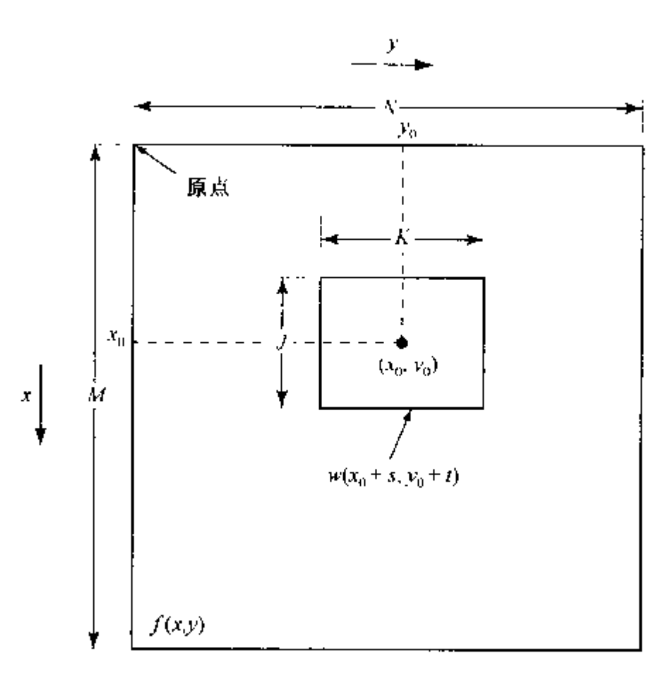


图 12.8　在点(x_0,y_0)得到 f 和 w 相关的排列

在式 (12.2.7) 中给出的相关函数存在着对于 f 和 w 的幅度变化过于敏感的缺点。例如,若 f 的所有值乘 2,$c(x,y)$的值也加倍。为克服这个缺点而经常使用的一种方法是,在匹配过程中使用相关系数,其定义如下:

$$\gamma(x,y)=\frac{\sum_s\sum_t[f(s,t)-\bar{f}(s,t)][w(x+s,y+t)-\bar{w}]}{\left\{\sum_s\sum_t[f(s,t)-\bar{f}(s,t)]^2\sum_s\sum_t[w(x+s,y+t)-\bar{w}]^2\right\}^{\frac{1}{2}}} \quad (12.2.8)$$

这里 $x=0,1,2,\cdots,M-1,y=0,1,2,\cdots N-1,\bar{w}$ 是 w 中的像素平均值(只计算一次),$\bar{f}$ 是 f 中与 w 当前所在位置相重合的区域平均值,总和的值通常由 f 和 w 两者的坐标代入后求得。相关系数 $\gamma(x,y)$ 在 -1 到 1 之间取值,与 f 和 w 幅值中的区间变化相独立(见习题 12.5)。

例 12.2 通过相关系数进行对象匹配

图 12.9 说明了刚才讨论过的概念。图 12.9(a)是 $f(x,y)$。图 12.9(b)是 $w(x,y)$。相关系数 $\gamma(x,y)$ 如图 12.9(c)所示。在 f 和 w 之间找到最佳匹配的地方,相关系数 $\gamma(x,y)$ 的值更大(更亮)。

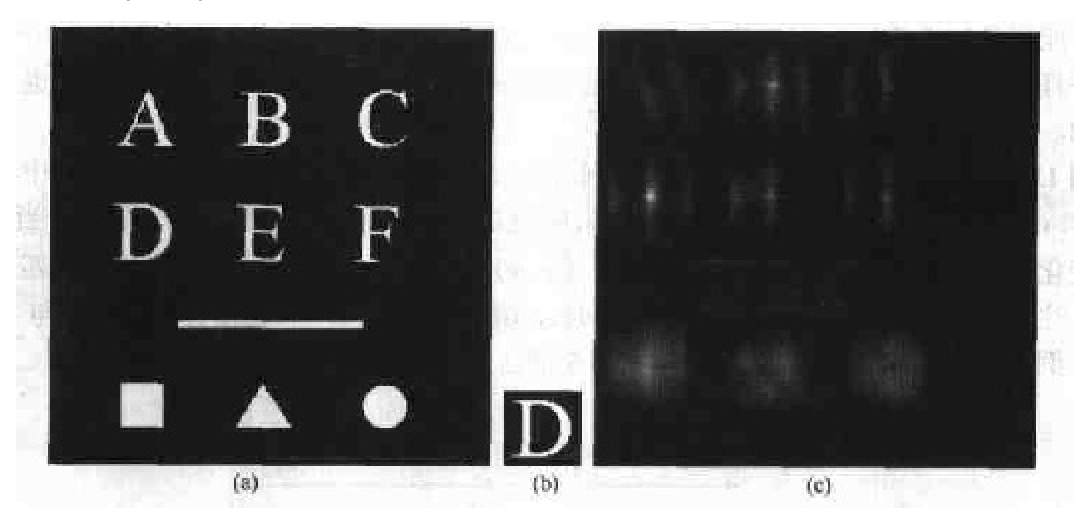


图 12.9 (a)图像,(b)子图,(c)(a)和(b)的相关系数。注意当子图(b)和(a)中的字母"D"一致时,(c)中出现最高值(亮点)

虽然相关函数对于幅度变化可以通过使用相关系数归一化,但要得到归一化的尺寸变化和旋转变化是困难的。尺寸的归一化涉及空间定标,这个过程本身会增加大量的计算。旋转变化的归一化更为困难。如果从 $f(x,y)$ 中可以提取其旋转变化方式的线索,就可以简单地对 $w(x,y)$ 进行旋转使它同 $f(x,y)$ 的旋转角度对准。然而,如果其旋转性质是未知的,那么寻找最佳匹配就要求对 $w(x,y)$ 进行全方位的旋转。这一过程是不实际的,当出现不确定的或不受约束的旋转变化时,相关很少使用。

在 4.6.4 节中我们曾提到,相关也可以通过 FFT 算法在频域内执行。如果 f 和 w 大小相同,这种方法比在空间域中直接实现图像的相关更有效。当 w 比 f 小得多时,可以使用式(12.2.7)。一种由 Campbell[1969]发明的折中估计法指出,如果 w 中的非零项数目小于 132(大约 13×13 像素的子图),则直接使用式(12.2.7)比 FFT 算法更为有效。当然,这个数目取决于使用的机器和算法,但是,它确实指出了在频域作为一种替代时应考虑子图的大致尺寸。在频域中使用相关系数更为困难。通常它是在空间域中直接计算出来的。

12.2.2 最佳统计分类器

在这一节中,我们将讨论一种用于模式识别的概率方法。在大多数处理测量和判断物理

事件的场合,对概率的考虑在模式识别中变得十分重要。因为,在这种情况下通常会产生随机的模式分类。正如下面讨论表明的那样,有可能推导出一种分类方法,这种方法在感觉上是最佳的,平均来说使用它产生分类错误的概率很低(见习题12.10)。

基础

特定模式 **x** 来自 ω_i 类的概率表示为 $p(\omega_i/\mathbf{x})$。如果模式分类器判断 **x** 来自类 ω_j,而实际上它来自类 ω_i,分类器就会出现一次失败的分类,表示为 L_{ij}。如果考虑模式 **x** 可能属于类 W 中的任何一种模式,则在将模式指定为类 ω_j 时的平均失效率为:

$$r_j(\mathbf{x}) = \sum_{k=1}^{W} L_{kj}p(\omega_k/\mathbf{x}) \tag{12.2.9}$$

这个公式在决策理论的术语上叫做条件平均风险或条件平均失效。

由基本概率理论我们知道 $p(A/B) = [p(A)p(B/A)]/p(B)$。应用此表达式,我们用下列形式书写式(12.2.9):

$$r_j(\mathbf{x}) = \frac{1}{p(\mathbf{x})}\sum_{k=1}^{W} L_{kj}p(\mathbf{x}/\omega_k)P(\omega_k) \tag{12.2.10}$$

这里 $p(\mathbf{x}/\omega_k)$是来自类 ω_k 的模式的概率密度函数,$P(\omega_k)$是类 ω_k 出现的概率。由于 $1/p(\mathbf{x})$ 是正的并且对所有的 $r_j(\mathbf{x}), j = 1,2,\cdots,W$ 都是如此,它可以从式(12.2.10)中被省略而不影响函数从最小值到最大值的相对顺序。平均失效率的表示就化简为:

$$r_j(\mathbf{x}) = \sum_{k=1}^{W} L_{kj}p(\mathbf{x}/\omega_k)P(\omega_k) \tag{12.2.11}$$

分类器有 W 个可能的类从任何已给出的未知模式进行选择。如果分类器对每一个模式 **x** 计算 $r_1(\mathbf{x}), r_2(\mathbf{x}),\cdots, r_W(\mathbf{x})$并且以最低的失效率对每一个模式制定相应的类,则对所有判决的总体平均失效率就会降至最低。这种能将总体平均失效率降至最低的分类器称做贝叶斯分类器。因此,贝叶斯分类器是在 $r_i(\mathbf{x}) < r_j(\mathbf{x}),\ j = 1,2,\cdots,W$ 且$j \neq i$ 时将一个未知模式 **x** 归属给类 ω_i 的。换句话说,**x** 是在下列公式成立时才归属类 ω_i 的。

$$\sum_{k=1}^{W} L_{ki}P(\mathbf{x}/\omega_k)P(\omega_k) < \sum_{q=1}^{W} L_{qj}p(\mathbf{x}/\omega_q)P(\omega_q) \tag{12.2.12}$$

对所有的 $j;j\neq i$。通常,对一个正确决策分类"失败"赋予0值,对任何不正确决策分类失败通常赋予一个相同的非零值(即赋1值)。在这样的条件下,失效函数的形式为:

$$L_{ij} = 1 - \delta_{ij} \tag{12.2.13}$$

这里当 $i = j$ 时 $\delta_{ij} = 1$,当 $i \neq j$ 时 $\delta_{ij} = 0$。式(12.2.13)表明对非正确决策的一个单位失效和对正确决策的一个0失效。将式(12.2.13)代入式(12.2.11)得到式:

$$\begin{aligned} r_j(\mathbf{x}) &= \sum_{k=1}^{W}(1 - \delta_{kj})p(\mathbf{x}/\omega_k)P(\omega_k) \\ &= p(\mathbf{x}) - p(\mathbf{x}/\omega_j)P(\omega_j) \end{aligned} \tag{12.2.14}$$

如果对所有 $j\neq i$ 时下列公式成立

$$p(\mathbf{x}) - p(\mathbf{x}/\omega_i)P(\omega_i) < p(\mathbf{x}) - p(\mathbf{x}/\omega_j)P(\omega_j) \tag{12.2.15}$$

或者等价地有如下公式成立时

$$p(\mathbf{x}/\omega_i)P(\omega_i) > p(\mathbf{x}/\omega_j)P(\omega_j) \quad j = 1,2,\cdots,W; j \neq i \tag{12.2.16}$$

则贝叶斯分类器将一个模式 $\mathbf{x}$ 赋予类 ω_i。

参考导出式(12.2.1)的讨论,可以了解到对于 0-1 失效函数的贝叶斯分类器不过是下列函数的决策值的计算:

$$d_j(\mathbf{x}) = p(\mathbf{x}/\omega_j)P(\omega_j) \quad j = 1,2,\cdots,W \tag{12.2.17}$$

这里哪个类的决策函数的值最大,模式矢量 $\mathbf{x}$ 就归属于哪一个类。

在式(12.2.7)中给出的判别函数在将错误分类的平均失效率降低到最低的能力方面是最佳的。然而,在每个类中的模式的概率密度函数和每种类出现的概率必须是已知的。后者的要求通常并不构成问题。比如,如果所有类的出现几率大致相同,就可令 $P(\omega_j) = 1/M$。即使这个条件不正确,我们也可以通过对问题的认识推断出这些概率。概率密度函数$p(\mathbf{x}/w_j)$的估计就是另一回事了。如果模式矢量 $\mathbf{x}$ 是 n 维的,那么 $p(\mathbf{x}/\omega_j)$ 就是一个 n 元函数,如果它的形式是未知的,就需要使用多元概率理论的方法对它进行估计。这类方法在实际使用中很难应用,尤其是代表每一个类的模式数目不大,或隐含的概率密度函数形式的规律性不佳时更是如此。由于这些原因,贝叶斯分类器的运用通常是基于以下假设:对不同概率密度函数的解析式及从每一类样本模式估计的必需参数。目前,对 $p(\mathbf{x}/\omega_j)$ 的假设形式普遍使用的是高斯概率密度函数。这种设定越接近真实情况,贝叶斯分类器方法在分类中越能接近最低平均失效率。

高斯模式类的贝叶斯分类器

首先,考虑一个包含两个模式类($W = 2$)的一维($n = 1$)问题,这两个模式类具有高斯密度,分别具有均值 m_1 和 m_2 与标准差 σ_1 和 σ_2。由式(12.2.17)可知贝叶斯判别函数具有如下形式:

$$\begin{aligned} d_j(x) &= p(x/\omega_j)P(\omega_j) \\ &= \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{(x-m_j)^2}{2\sigma_j^2}} P(\omega_j) \quad ,j = 1,2 \end{aligned} \tag{12.2.18}$$

这里,模式此时是标量,用 x 表示。图 12.10 显示了两个类概率密度函数的曲线图。两个类之间的交界是一个单点,用 x_0 表示,使得 $d_1(x_o) = d_2(x_0)$。如果两个类出现几率相等,则 $P(\omega_1) = P(\omega_2) = 1/2$,并且决策边界就是 $p(x_0/\omega_1) = P(x_0/\omega_2)$处 x_0 的值。这一点如图 12.10所示,是两个概率密度函数的交集。任何处于 x_0 点右侧的模式(点)都被划归类 ω_1。同样,任何处于 x_0 点左侧的模式(点)都被划归类 ω_2。当两个类出现几率不相等时,如果类 ω_1 出现几率大则 x_0 点左移,或者相反,如果类 ω_2 出现几率大则 x_0 点右移。这就是所期望的结果,因为分类器总是试图将错误分类的失效降至最低。例如,在极端情况下,如果类 ω_2 从不出现,通过总将所有模式划归类 ω_1,分类器也不会出现错误(即点 x_0 趋于负无穷)。

在 n 维情况下,第 j 个模式类中矢量的高斯密度具有如下所示的形式:

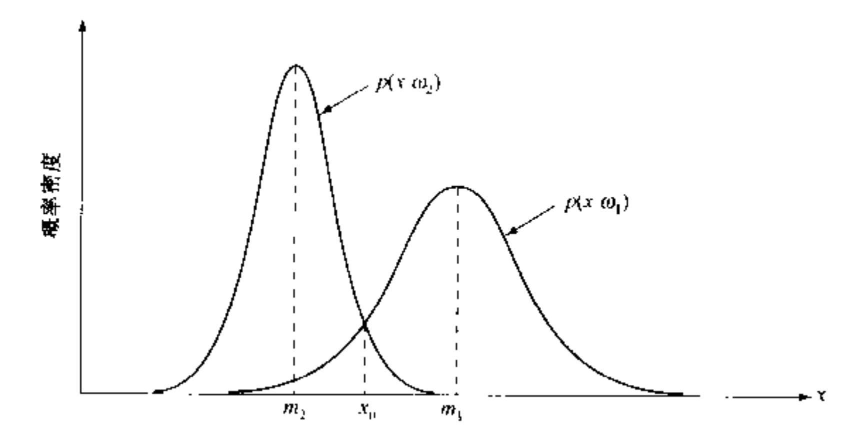


图 12.10 两个一维模式类的概率密度函数。如果两个类出现几率相等,则点 x_0 就是决策边界

$$p(\mathbf{x}/\omega_j) = \frac{1}{(2\pi)^{n/2}|\mathbf{C}_j|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m}_j)^T\mathbf{C}_j^{-1}(\mathbf{x}-\mathbf{m}_j)} \tag{12.2.19}$$

这里每一个密度函数都由其矢量平均数 $\mathbf{m}_j$ 和协方差矩阵 $\mathbf{C}_j$ 完全指定,它被定义如下:

$$\mathbf{m}_j = E_j\{\mathbf{x}\} \tag{12.2.20}$$

和

$$\mathbf{C}_j = E_j\{(\mathbf{x} - \mathbf{m}_j)(\mathbf{x} - \mathbf{m}_j)^T\} \tag{12.2.21}$$

这里 $E_j\{\cdot\}$ 代表类 ω_j 中模式的变元期望值。在式(12.2.19)中,n 是模式矢量的维度,$|\mathbf{C}_j|$是矩阵 $\mathbf{C}_j$ 的行列式。用问题中量值的平均值近似期望值 E_j,得到平均矢量值的估计值和协方差矩阵如下:

$$\mathbf{m}_j = \frac{1}{N_j}\sum_{\mathbf{x}\in\omega_j}\mathbf{x} \tag{12.2.22}$$

和

$$\mathbf{C}_j = \frac{1}{N_j}\sum_{\mathbf{x}\in\omega_j}\mathbf{x}\mathbf{x}^T - \mathbf{m}_j\mathbf{m}_j^T \tag{12.2.23}$$

这里 N_j 是来自类 ω_j 的模式矢量数目,并且总和取全部向量。在这一节的后边,将给出一个例子示范如何使用这两个表达式。

协方差矩阵是对称的并且是半正定的。如 11.4 节中解释的那样,对角线元素 c_{kk} 是模式矢量第 k 个元素的方差。非对角线元素 c_{jk} 是 x_j 和 x_k 的协方差。多元高斯密度函数在协方差矩阵的非对角线元素均为零时还原为 $\mathbf{x}$ 每一个元素的单变元高斯密度的乘积。这种情况是在矢量变元 x_j 和 x_k 不相关时出现的。

依据式(12.2.17),类 ω_j 的贝叶斯判别函数为 $d_j(\mathbf{x}) = p(\mathbf{x}/\omega_j)P(\omega_j)$。然而,由于高斯密度函数的指数形式,用这个判别函数的自然对数形式更为方便。换句话说,可以使用下列形式的公式:

$$\begin{aligned} d_j(\mathbf{x}) &= \ln[p(\mathbf{x}/\omega_j)P(\omega_j)] \\ &= \ln p(\mathbf{x}/\omega_j) + \ln P(\omega_j) \end{aligned} \tag{12.2.24}$$

从分类性能的角度来说这个表达式同式(12.2.17)是等价的。因为对数是一个单调递增函数。换句话说,式(12.2.17)中的判别函数的顺序和式(12.2.24)中的顺序是一样的。将式(12.2.19)代入式(12.2.24)得到式(12.2.25)。

$$d_j(\mathbf{x}) = \ln P(\omega_j) - \frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\mathbf{C}_j| - \frac{1}{2}[(\mathbf{x}-\mathbf{m}_j)^T\mathbf{C}_j^{-1}(\mathbf{x}-\mathbf{m}_j)] \qquad (12.2.25)$$

项$(n/2)\ln 2\pi$对所有的类都是相同的,因此,可以从式(12.2.25)中消去。成为下列形式:

$$d_j(\mathbf{x}) = \ln P(\omega_j) - \frac{1}{2}\ln|\mathbf{C}_j| - \frac{1}{2}[(\mathbf{x}-\mathbf{m}_j)^T\mathbf{C}_j^{-1}(\mathbf{x}-\mathbf{m}_j)] \qquad (12.2.26)$$

对于$j = 1,2,\cdots,W$。式(12.2.26)表示高斯模式类在0-1失效函数条件下的贝叶斯判别函数。

式(12.2.26)中的判别函数是超二次曲面(n维空间中的二次函数),因为式子中出现的$\mathbf{x}$项没有高于二次的。毫无疑问,一个高斯模式的贝叶斯分类器最多做到在每对模式类之间置一个二阶判别曲面。不过,如果模式域确实是高斯的,则在分类中不会有其他曲面所产生的失效率少于平均失效率。

如果所有的协方差矩阵都相等,即$\mathbf{C}_j = \mathbf{C}, j = 1,2,\cdots,W$。通过展开式(12.2.26)并消去所有独立于$j$的项,会得到下列公式:

$$d_j(\mathbf{x}) = \ln P(\omega_j) + \mathbf{x}^T\mathbf{C}^{-1}\mathbf{m}_j - \frac{1}{2}\mathbf{m}_j^T\mathbf{C}^{-1}\mathbf{m}_j \qquad (12.2.27)$$

这是一个线性判别函数(超平面),$j = 1,2,\cdots,W$.

另外,如果$\mathbf{C}=\mathbf{I}$,这里$\mathbf{I}$代表单位距阵,且$P(\omega_j) = 1/W, j = 1,2,\cdots,W$,则有公式:

$$d_j(\mathbf{x}) = \mathbf{x}^T\mathbf{m}_j - \frac{1}{2}\mathbf{m}_j^T\mathbf{m}_j, j = 1,2,\cdots,W \qquad (12.2.28)$$

这些公式是最小距离分类器的判别函数,如式(12.2.5)所给出的一样。如果(1)模式类都是高斯的,(2)所有协方差矩阵等于单位矩阵,并且(3)所有类出现的几率相等,则最小距离分类器在贝叶斯意义上是最佳的。满足这些条件的高斯模式类是n维空间中外形相同的球状云团(称为超球面)。最小距离分类器在每对类之间设置一个超平面,这个超平面的特性是垂直等分连接每对超球面中心的线段。在两个维度上,类组成圆形区域,并且边界变为等分连接每对圆环中心的线段的直线。

例 12.3 三维模式的贝叶斯分类器

图12.11显示了一个由三维空间中的两个类组成的简单排列。我们用这些模式说明运用贝叶斯分类器的机理,假设每个类中的模式都是服从高斯分布的样本。

对图12.11中的模式应用式(12.2.22)得到:

$$\mathbf{m}_1 = \frac{1}{4}\begin{bmatrix}3\\1\\1\end{bmatrix} \quad 和 \quad \mathbf{m}_2 = \frac{1}{4}\begin{bmatrix}1\\3\\3\end{bmatrix}$$

类似地,对两个模式类依次使用式(12.2.23)得到两个协方差矩阵,在这种情况下两个矩阵是相等的:

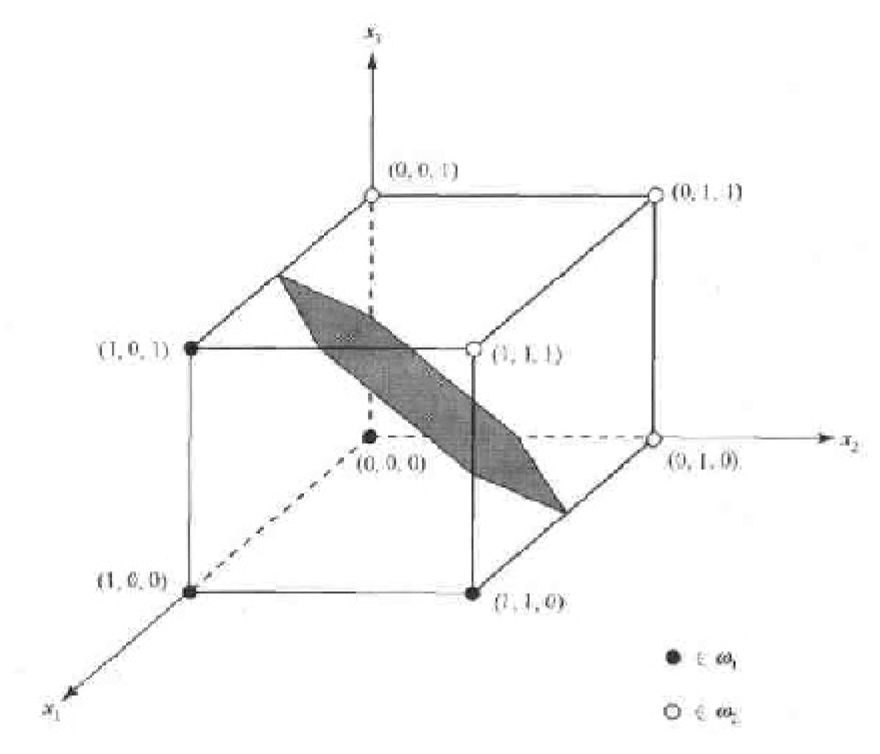


图 12.11　两个简单模式类和它们的贝叶斯判别边界(阴影处)

$$\mathbf{C}_1 = \mathbf{C}_2 = \frac{1}{16}\begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix}$$

因为协方差矩阵相等，则贝叶斯判别函数由式(12.2.27)给出。如果假定 $P(\omega_1) = P(\omega_2) = 1/2$，然后应用式(12.2.28)会得到：

$$d_j(\mathbf{x}) = \mathbf{x}^T C^{-1}\mathbf{m}_j - \frac{1}{2}\mathbf{m}_j^T \mathbf{C}^{-1}\mathbf{m}_j$$

这里

$$\mathbf{C}^{-1} = \begin{bmatrix} 8 & -4 & -4 \\ -4 & 8 & 4 \\ -4 & 4 & 8 \end{bmatrix}$$

对 $d_j(\mathbf{x})$执行矢量矩阵展开，规定判别函数：

$$d_1(\mathbf{x}) = 4x_1 - 1.5 \quad 和 \quad d_2(\mathbf{x}) = -4x_1 + 8x_2 + 8x_3 - 5.5$$

分开两个类的决策面是：

$$d_1(\mathbf{x}) - d_2(\mathbf{x}) = 8x_1 - 8x_2 - 8x_3 + 4 = 0$$

图 12.11 显示了这个面的一部分，注意到类被有效地分开了。

贝叶斯分类器方法最成功的应用之一是，对在航天器、卫星或空间站上利用多频谱扫描器产生的遥感图像进行分类处理。对这些平台上产生的大量图像数据进行自动图像分类和分析处理，是在遥感领域相当有意义的一项任务。遥感技术应用于许多领域，包括在陆地上的应

用,如农作物产量的调查,农作物灾害检测,森林、空气质量和水质的监测,地质研究,天气预报以及许多对环境有重要意义的领域。下面的例子显示了一个典型应用。

例 12.4 使用贝叶斯分类器对多频谱数据的分类

如在 1.3.4 节和 11.4 节曾讨论过的,多频谱扫描器在选定的频段对电磁能量有响应,例如 0.40 ~ 0.44 μm,0.58 ~ 0.62 μm,0.66 ~ 0.72 μm 和 0.80 ~ 1.00 μm 波段。这些频率范围分别处于紫色光、绿色光、红色光和红外频段上,在这些频段上对一个区域进行扫描,生成 4 幅数字图像,每个频段产生一幅。如果对这些图像进行配准(实际上经常是这样做的),就可以将它们前后叠加在一起形成一幅图像呈现出来,如图 12.12 所示。因而,就如同在 11.4 节中做过的,地面上的每一个点都可以采用 $\mathbf{x} = (x_1, x_2, x_3, x_4)^T$ 表示成一个四元矢量的形式。这里 x_1 代表紫外光的像,x_2 代表绿光的像,等等。如果图像是 512 × 512 像素大小的,每幅四频段的叠加图可以用 262 144 个四维模式矢量表示。

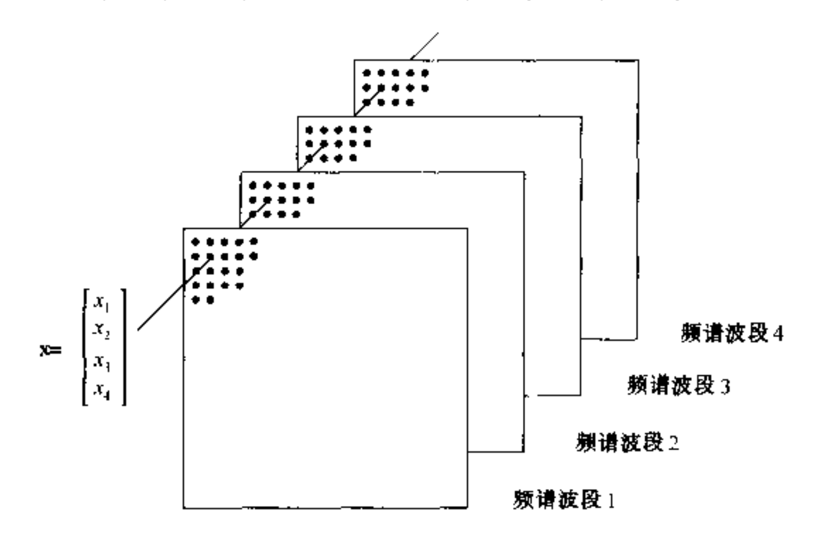
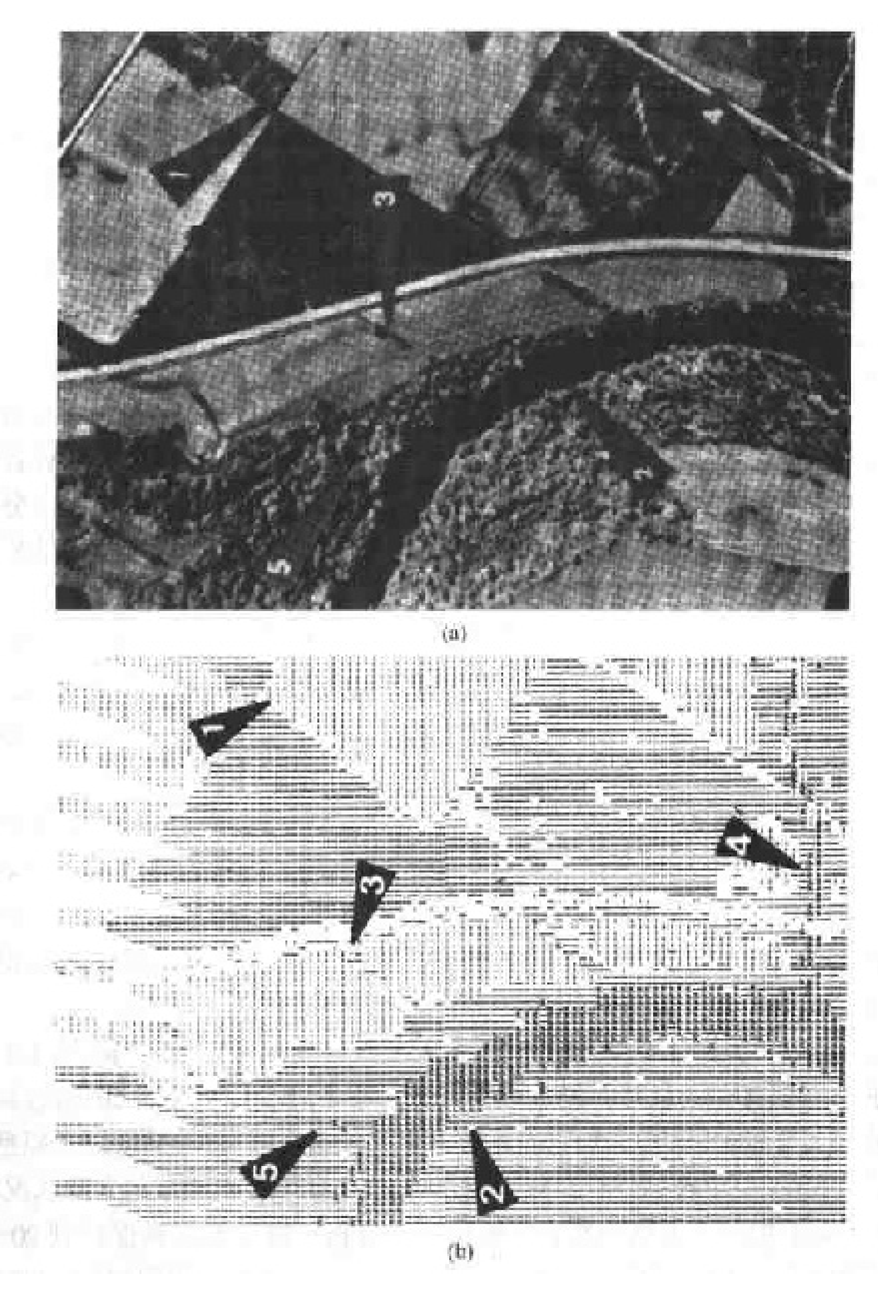


图 12.12　由多频谱扫描器生成的 4 幅数据图像经像素配准后的模式矢量格式

如以前所注意到的,高斯模式的贝叶斯分类器需要对每类平均矢量和协方差矩阵的估计。在遥感应用中,这些估计是通过像前面的范例中提到的那样,收集感兴趣地区的多频谱数据并运用这些样本而得到的。图 12.13(a)显示的是一幅从航天器上得到的典型遥感图像(本图是由一幅多频谱原图转换成的单色图)。在这种特殊情况下,问题是将所拍区域划分为植被、水和裸露的土地等不同区域。图 12.13(b)显示的是用高斯型贝叶斯分类器分类的结果。箭头显示了我们感兴趣的特征。箭头 1 指向一块绿色植被区的拐角处。箭头 2 指出了一条河。箭头 3 指向两块裸露土地之间的灌木树篱。箭头 4 指出了被系统正确识别出来的一条支流。箭头 5 指出了在图 12.13(a)中几乎无法分辨的一个小池塘。把原始图像与计算机输出的识别结果相比较可以看出,它与人类通过视觉分析得到的识别结果非常相似。

在结束这一节之前,有意思地注意到,逐点分类方法确实可以将图像分成不同的类,正如我们在前述例子中描述的那样。这种方法类似于用多个变量的门限分段法,这种方法在 10.3.7 节中有简要的描述。



(a)

(b)

图12.13 (a)多频谱图像,(b)打印出来的使用了贝叶斯分类器的机器分类结果(由Purdue大学遥感应用实验室提供)

12.2.3 神经网络

这种方法在以前的两节中有所论述。它基于使用样本模式来估计每个模式类的统计参数。最小距离分类器完全由每个类的平均矢量确定。同样,具有高斯密度的贝叶斯分类器完全由每个类的平均矢量和协方差矩阵确定。用于估计这些参数的模式(已知其所属的类)称做训练模式,一系列从每个类中得到的这类模式称做一个训练集合。使用训练集合得到判别函数的过程叫做学习或训练。

在刚刚讨论过的两种方法中,训练是简单的。每一个类的训练模式用于计算符合本类的判别函数参数。在估计了问题中的参数后,分类器的结构就被固定下来了,而且分类器的最终

效果的好坏取决于真实的模式类密度是否满足基本的统计假设。此处使用的统计假设来自所使用的分类方法。

对一个要解决的问题来说,模式类的统计特性通常是未知的或者无法估计的(回想我们在前面章节谈到的应用多元统计所遇到的困难)。实际上,这类决策理论问题最好使用直接通过训练过程生成所需判别函数这一方法来处理。对基本的概率密度函数或关于模式类的其他概率信息进行假设是必要的。在这一节中,我们将讨论各种符合这一准则的方法。

背景

以下所列材料的精髓就是使用大量非线性元素的计算单元(称为神经元),这些神经元是以被认为同大脑神经元的互联方式一样的方法组织起来的。其产生的模型有各种各样的名字,包括神经网络、神经计算机、分布式并行处理(PDP)模型、神经形态学系统、分层次自适应网络和连接模式。在这里,使用"神经网络"这一名称或简称"神经网"。使用这些网络作为工具,通过对模式训练集合的进一步描述自适应地导出判别函数系数。

神经网络受到关注可以追溯到20世纪40年代早期McCulloch和Pitts [1943]的示范性工作。他们提议用二进制门限器件和包括0-1和1-0的神经状态变化的随机算法作为神经系统建模的基础。由Hebb [1949]进行的后续工作是建立在数学模型基础上的,这个数学模型试图通过增强或建立关联来掌握"学习"这一概念。

20世纪50年代中期至20世纪60年代早期,由Rosenblatt [1959,1962]发明的一种称做学习机的装置在模式识别理论领域的研究者和开创者中引起了巨大的轰动。人们对这种称做感知器的机器给予如此关注的原因是,由于数学的发展,证明这种感知器在使用线性离散训练集合(也就是可被超平面分开的训练集合)进行训练后,可以通过有限的迭代步骤得出解答。这个解答采取超平面系数(它们能正确分离由训练集模式描述的类)的形式。

遗憾的是,在接下来研究学习模式的过程中,人们很快便大失所望。当时基本的感知器和它的一些推广对于大部分有实际意义的模式识别任务是很不够的。后来,试图通过将这种装置多层化来扩展类感知器装置的能力,尽管它们在概念上都较有吸引力,但是均缺乏原型感知器那样有效的训练算法。Nilsson [1965]总结了20世纪70年代中期学习机领域的大致状况。几年之后,Minsky和Papert [1969]提出了类感知器装置的局限性分析。这个看法被保持到20世纪80年代中期。这一观点在Simon[1986]的评论中得到了证明。在这一工作中,原始论文1984年以法文出版,在这篇文章中,Simon以"一个神话的诞生和死亡"为题目对感知器进行驳斥。

由Rumelhart,Hinton和Williams[1986]在多重感知器新的训练运算法则的发展中得出的更新结果,使这个问题发生了很大变化。他们的基本方法,即被称做反向传播方式学习的一般性德尔塔(delta)规则,为多层机器提供了一种有效的训练方法。尽管这种训练算法与单层感知器比起来,在相似的证明方面无法得出一个集中的解,但产生的德尔塔规则在解决大量实际问题时却得到了成功的应用。这一成功奠定了多层类感知器装置作为现今神经网络应用中的主要模型之一的地位。

两个模式类的感知器

在这种最基本的形式中,感知器"学习"了一个线性判别函数,这个判别函数将两个线性分离的训练集合对半分开。图12.14(a)示意性地显示了两个模式类的感知器模型。

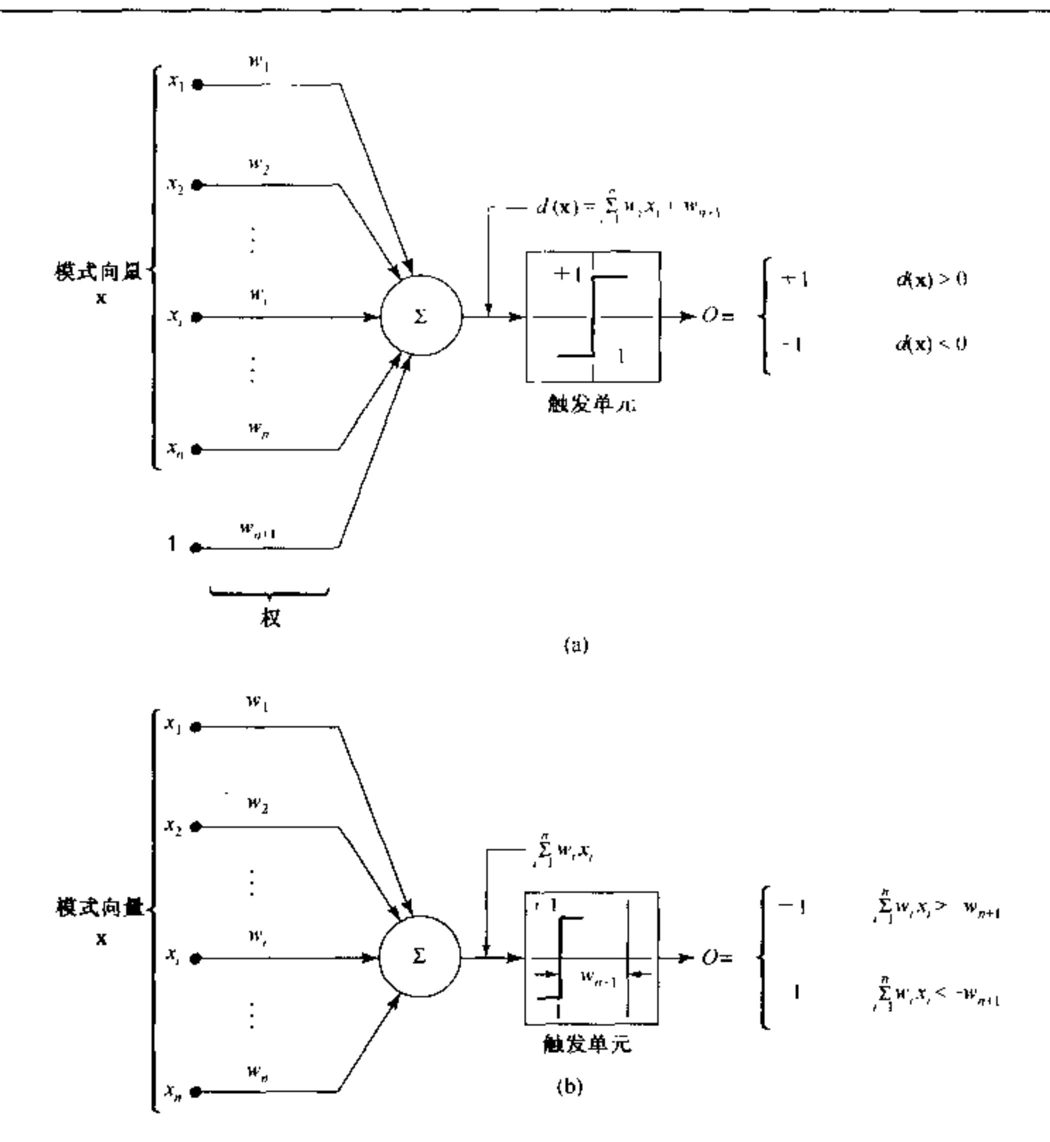


图 12.14 对两个模式类的感知器模型的等效描述

这个基本装置的响应是基于输入值的加权和,即:

$$d(\mathbf{x}) = \sum_{i=1}^{n} w_i x_i + w_{n+1} \tag{12.2.29}$$

这是一个考虑了模式矢量的分量的线性判别函数。系数 $w_i, i = 1,2,\cdots,n,n+1$,称为权,在元素相加并被输入到门限元件中之前对输入进行修正。在这方面,权值同人类大脑神经系统中的神经突触是相似的。把总连接的输出映射到最终的装置输出的函数有时称为激活函数。

当 $d(\mathbf{x}) > 0$ 时门限元件使感知器的输出为 +1,表示模式 $\mathbf{x}$ 被识别并划归类 ω_1。当 $d(\mathbf{x}) < 0$时相反的情况也是正确的。这种操作模式与前面做出的和式(12.2.2)相关的评论是一致的,那时考虑对两个模式类使用单个判别函数进行判别。当 $d(\mathbf{x}) = 0$ 时,$\mathbf{x}$ 位于分离两个模式类的判别(决策)平面上,给出了一个不能确定的条件。由感知器实现的判别边界是通过设式(12.2.29)等于 0 得到的。

$$d(\mathbf{x}) = \sum_{i=1}^{n} w_i x_i + w_{n+1} = 0 \tag{12.2.30}$$

或

$$w_1x_1 + w_2x_2 + \cdots + w_nx_n + w_{n+1} = 0 \tag{12.2.31}$$

这是在 n 维模式空间中的一个超平面方程。在几何上,前 n 个系数确定了超平面的方位,而末尾的系数是与从原点到超平面的垂直距离成比例的。因此,如果 $w_{n+1}=0$,超平面通过模式空间的原点。同样,如果 $w_j=0$,超平面平行于 x_j 轴。

图 12.14(a)中门限元素的输出取决于 $d(\mathbf{x})$的符号。代替测试全部函数以确定它是正还是负,可以对项 w_{n+1}检验式(12.2.29)的和,此时系统的输出是:

$$O = \begin{cases} +1 & \sum_{i=1}^{n} w_ix_i > -w_{n+1} \\ -1 & \sum_{i=1}^{n} w_ix_i < -w_{n+1} \end{cases} \tag{12.2.32}$$

在图 12.14(b)中所示的实现方案同图 12.14(a)中所示的等价,仅有的区别是,门限函数表示成了量 $-w_{n+1}$,并且不存在常量 1 的输入。当在这一节的后面讨论多层神经网络的实现时,再回顾这两个表达形式的等价性。

另一个经常使用的公式是,通过增加第$(n+1)$个元素对模式矢量进行扩展,这个元素总等于 1,而与类成员无关。即,扩展的模式矢量 $\mathbf{y}$ 是从一个模式矢量 $\mathbf{x}$ 通过令 $y_i = x_i, i=1,2,\cdots,n$,并增加元素 $y_{n+1}=1$ 得到的。式(12.2.29) 就变成:

$$\begin{aligned} d(\mathbf{y}) &= \sum_{i=1}^{n+1} w_iy_i \\ &= \mathbf{w}^T\mathbf{y} \end{aligned} \tag{12.2.33}$$

这里 $\mathbf{y}=(y_1,y_2,\cdots,y_n,1)^T$ 现在是一个扩展模式矢量,并且 $\mathbf{w}=(w_1,w_2,\cdots,w_n,w_{n+1})^T$ 称为权矢量。这个表达式通常在表达符号方面更为简便。然而忽略使用的公式,关键问题是使用一个给定模式矢量的训练集合找到 $\mathbf{w}$,这个模式矢量训练集合来自两类中的一个。

训练算法

下面展开讨论的算法都是过去这些年来为训练感知器提出的大量方法中具有代表性的方法。

线性可分离的类

对于两个线性可分离的训练集为了得到权重矢量可遵循一个简单的迭代算法。对于两个分别属于类 ω_1 和 ω_2 的扩展了模式矢量的训练集合,令 $\mathbf{w}(1)$代表任选的一个类初始权重矢量。然后,在迭代的第 k 步,如果 $\mathbf{y}(k)\in\omega_1$ 并且 $\mathbf{w}^T(k)\mathbf{y}(k) \leqslant 0$,则用下列公式代替 $\mathbf{w}(k)$:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + c\mathbf{y}(k) \tag{12.2.34}$$

这里 c 是一个正的补偿增量。相反,如果 $\mathbf{y}(k)\leqslant \omega_2$ 并且 $\mathbf{w}^T(k)\mathbf{y}(k) \geqslant 0$,则用下列公式代替 $\mathbf{w}(k)$:

$$\mathbf{w}(k+1) = \mathbf{w}(k) - c\mathbf{y}(k) \tag{12.2.35}$$

否则,令 $\mathbf{w}(k)$不变:

$$\mathbf{w}(k+1)=\mathbf{w}(k) \tag{12.2.36}$$

这个算法只有模式在训练过程的第 k 步被错误地分类时才改变 $\mathbf{w}$。假设校正增量 c 是正的，现在来说是一个常量。有时候这个算法作为固定增量校正规则。

当两个类的整个训练集通过机器循环而没有出现任何错误时，算法就会收敛。如果模式的两个训练集合是线性可分离的，那么固定增量校正规则在有限步内收敛。这个结果的证明称做感知器训练定理，可以在 Duda，Hart 和 Stork [2001]，Tou 和 Gonzalez [1974] 和 Nilsson [1965]等人所著的书中找到。

例 12.5　*感知器算法解释*

考虑图 12.15(a)所示的两个训练集，每一个集由两个模式组成。因为两个训练集是线性可分离的，所以可以成功应用训练算法。算法应用之前，模式先被扩展。对类 ω_1 生成训练集合 $\{(0,0,1)^T,(0,1,1)^T\}$，对 ω_2 生成训练集合 $\{(1,0,1)^T,(1,1,1)^T\}$。令 $c=1$，$\mathbf{w}(1)=0$，并按下列顺序生成模式：

$$\mathbf{w}^T(1)\mathbf{y}(1)=[0,0,0]\begin{bmatrix}0\\0\\1\end{bmatrix}=0 \qquad \mathbf{w}(2)=\mathbf{w}(1)+\mathbf{y}(1)=\begin{bmatrix}0\\0\\1\end{bmatrix}$$

$$\mathbf{w}^T(2)\mathbf{y}(2)=[0,0,1]\begin{bmatrix}0\\1\\1\end{bmatrix}=1 \qquad \mathbf{w}(3)=\mathbf{w}(2)=\begin{bmatrix}0\\0\\1\end{bmatrix}$$

$$\mathbf{w}^T(3)\mathbf{y}(3)=[0,0,1]\begin{bmatrix}1\\0\\1\end{bmatrix}=1 \qquad \mathbf{w}(4)=\mathbf{w}(3)-\mathbf{y}(3)=\begin{bmatrix}-1\\0\\0\end{bmatrix}$$

$$\mathbf{w}^T(4)\mathbf{y}(4)=[-1,0,0]\begin{bmatrix}1\\1\\1\end{bmatrix}=-1 \qquad \mathbf{w}(5)=\mathbf{w}(4)=\begin{bmatrix}-1\\0\\0\end{bmatrix}$$

这里，如式(12.2.34)和式(12.2.35)所示，因为分类错误，权重矢量中的校正在第 1 步和第 3 步中做出。因为只有当通过所有训练模式，算法产生一个完全无错误的迭代时，才可能得到一个解，所以训练集必须被多次提供。机器的学习过程是不断以同样的方式令 $\mathbf{y}(5)=\mathbf{y}(1)$，$\mathbf{y}(6)=\mathbf{y}(2)$，$\mathbf{y}(7)=\mathbf{y}(3)$，$\mathbf{y}(8)=\mathbf{y}(4)$。$k=14$ 时收敛，生成权值矢量解 $\mathbf{w}(14)=(-2,0,1)^T$。相应的判别函数是 $d(\mathbf{y})=-2y_1+1$。通过令 $x_i=y_i$ 返回初始的模式空间，生成 $d(\mathbf{x})=-2x_1+1$。当集合等于 0 时，判别函数变成图 12.15(b)中所示的判别边界的方程式。

不可分离的类

实际上，线性可分离模式类是个例外，不常见到。因此，在 20 世纪 60 年代和 20 世纪 70 年代，大量有意义的研究工作转向处理不可分离的模式类的技术开发领域。伴随着最近在神经网络训练方面取得的进展，许多处理不可分离模式的方法成为只具有历史意义的项目。然而，早期的方法之一与这一讨论有直接的关系，这就是原始的德尔塔规则。用于训练感知器的众所周知的 Widrow-Hoff[或叫做最小均方(LMS)]德尔塔规则，在训练的每一步都可以在真实的和期望

的响应间将错误降至最低。

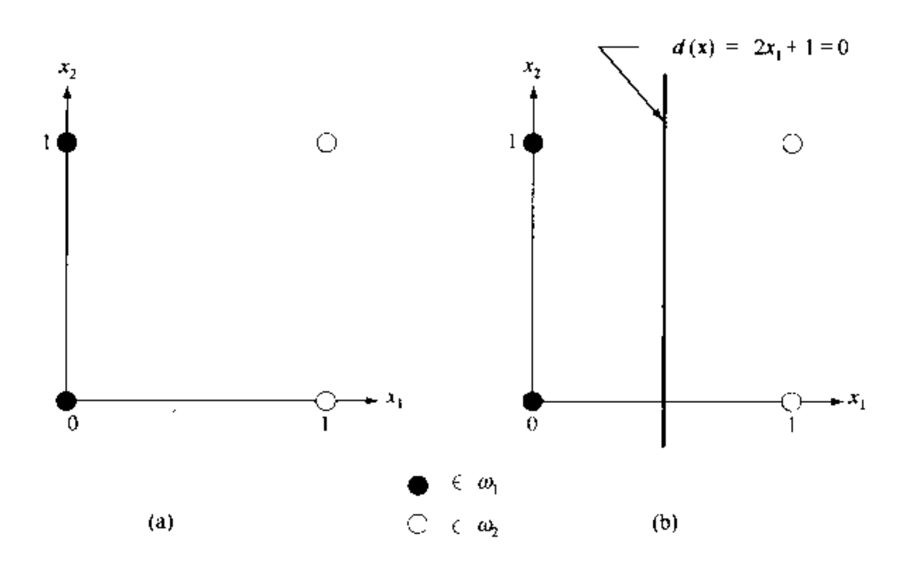


图 12.15 (a)属于两个类的模式,(b)通过训练确定的判别边界

考虑准则函数:

$$J(\mathbf{w}) = \frac{1}{2}(r - \mathbf{w}^T\mathbf{y})^2 \tag{12.2.37}$$

这里 r 是期望的响应(即,当扩展训练模式矢量 y 属于类 ω_1 时,$r = +1$,当 y 属于类 ω_2 时,$r = -1$)。目的是当 $r = \mathbf{w}^T\mathbf{y}$ 时,在 $J(\mathbf{w})$ 梯度的相反方向逐渐调整 $\mathbf{w}$,以便找到函数的最小值,即,对应正确分类的最小值。如果 $\mathbf{w}(k)$ 表示迭代的第 k 步的权重矢量,则梯度下降算法可写成下式:

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \alpha\left[\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}}\right]_{\mathbf{w}=\mathbf{w}(k)} \tag{12.2.38}$$

这里 $\mathbf{w}(k + 1)$是 $\mathbf{w}$ 的新值,并且 $\alpha > 0$ 给出校正量。由式(12.2.37)得到:

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = -(r - \mathbf{w}^T\mathbf{y})\mathbf{y} \tag{12.2.39}$$

把这一结果代入式(12.2.38)得到:

$$\mathbf{w}(k + 1) = \mathbf{w}(k) + \alpha[r(k) - \mathbf{W}^T(k)\mathbf{y}(k)]\mathbf{y}(k) \tag{12.2.40}$$

所用的初始权重矢量 $\mathbf{w}(1)$是任意的。

在权重矢量中,定义增量为:

$$\Delta\mathbf{w} = \mathbf{w}(k + 1) - \mathbf{w}(k) \tag{12.2.41}$$

以德尔塔校正算法的形式写出式(12.2.40):

$$\Delta\mathbf{w} = \alpha e(k)\mathbf{y}(k) \tag{12.2.42}$$

这里

$$e(k) = r(k) - \mathbf{w}^T(k)\mathbf{y}(k) \tag{12.2.43}$$

当模式 $\mathbf{y}(k)$存在时,$e(k)$是涉及权重矢量 $\mathbf{w}(k)$的误差。

式(12.2.43)给出了权重矢量 $\mathbf{w}(k)$的误差。如果把它变成 $\mathbf{w}(k + 1)$,但保留模式不变,

误差就变成：

$$e(k) = r(k) - \mathbf{w}^T(k+1)\mathbf{y}(k) \tag{12.2.44}$$

误差的变化量是：

$$\begin{aligned}\Delta e(k) &= [r(k) - \mathbf{w}^T(k+1)\mathbf{y}(k)] - [r(k) - \mathbf{w}^T(k)\mathbf{y}(k)] \\ &= -[\mathbf{w}^T(k+1) - \mathbf{w}^T(k)]\mathbf{y}(k) \\ &= -\Delta\mathbf{w}^T\mathbf{y}(k)\end{aligned} \tag{12.2.45}$$

但，$\Delta\mathbf{w} = \alpha e(k)\mathbf{y}(k)$，所以：

$$\Delta e = -\alpha e(k)\mathbf{y}^T(k)\mathbf{y}(k) = -\alpha e(k)\|\mathbf{y}(k)\|^2 \tag{12.2.46}$$

因此，权重的变化由于有因子 $\alpha\|\mathbf{y}(k)\|^2$ 而减少了误差。下一个输入模式开始新的自适应循环过程，使用因子 $\alpha\|\mathbf{y}(k+1)\|^2$ 减少下一个误差，如此继续下去。

α 的选择控制着稳定性和收敛速度(Widrow 和 Stearns [1985])。稳定性要求 $0 < \alpha < 2$。α 的实际范围在 $0.1 < \alpha < 1.0$ 之间。尽管这里没有给出证明，但实际上式(12.2.40)或式(12.2.42)和式(12.2.43)收敛于模式训练集合上一个最小均方误差解。当模式类可分离的时候，使用刚才讨论的算法得出的解也许会产生一个分离超平面。就是说，从感知器训练理论的角度看，一个均方误差解并不意味着会真有一个解。这种不确定性是在这个特殊的公式中在可分离和不可分离的情况下使用收敛算法所付出的代价。

到目前为止讨论过的这两个感知器训练算法可以推广到两类以上和非线性判别函数领域。基于先前做出的历史评价，寻找多类训练算法没有太大价值。相反，将在神经网络的内容中讨论多类训练问题。

多层前馈神经网络

在这一节中，主要关注多类模式识别问题的判别函数，包括由感知器计算节点层构成的体系结构，而不考虑类是否是可分离的。

基本的体系结构

图 12.16 显示了考虑的神经网络模型的体系结构。体系由排列结构相同的计算节点(神经元)层构成。这样，一层上每个神经元的输出都输入到下一层的每个神经元。称为 A 的第一层中的神经元个数为 N_A。一般 $N_A = n$，是输入的模式矢量维度。称为 Q 层的输出层的神经元数目由 N_Q 代表。N_Q 等于 W，是已经训练好用来识别的神经网络模式类数目。正如下面进行讨论的那样，如果网络的第 i 个输出为“高”，而其他输出为“低”，则网络识别模式矢量 $\mathbf{x}$ 属于类 ω_i。

如图 12.16 中的放大显示图所示，每个神经元都有与先前讨论的感知模型相同的形式(见图 12.14)，除了硬限激活函数被软限“S”函数代替。沿着神经网络所有路径的可微分性也是推导训练规则所需要的。下列的“S”激活函数具有需要的微分性：

$$h_j(I_j) = \frac{1}{1 + e^{-(I_j + \theta_j)/\theta_o}} \tag{12.2.47}$$

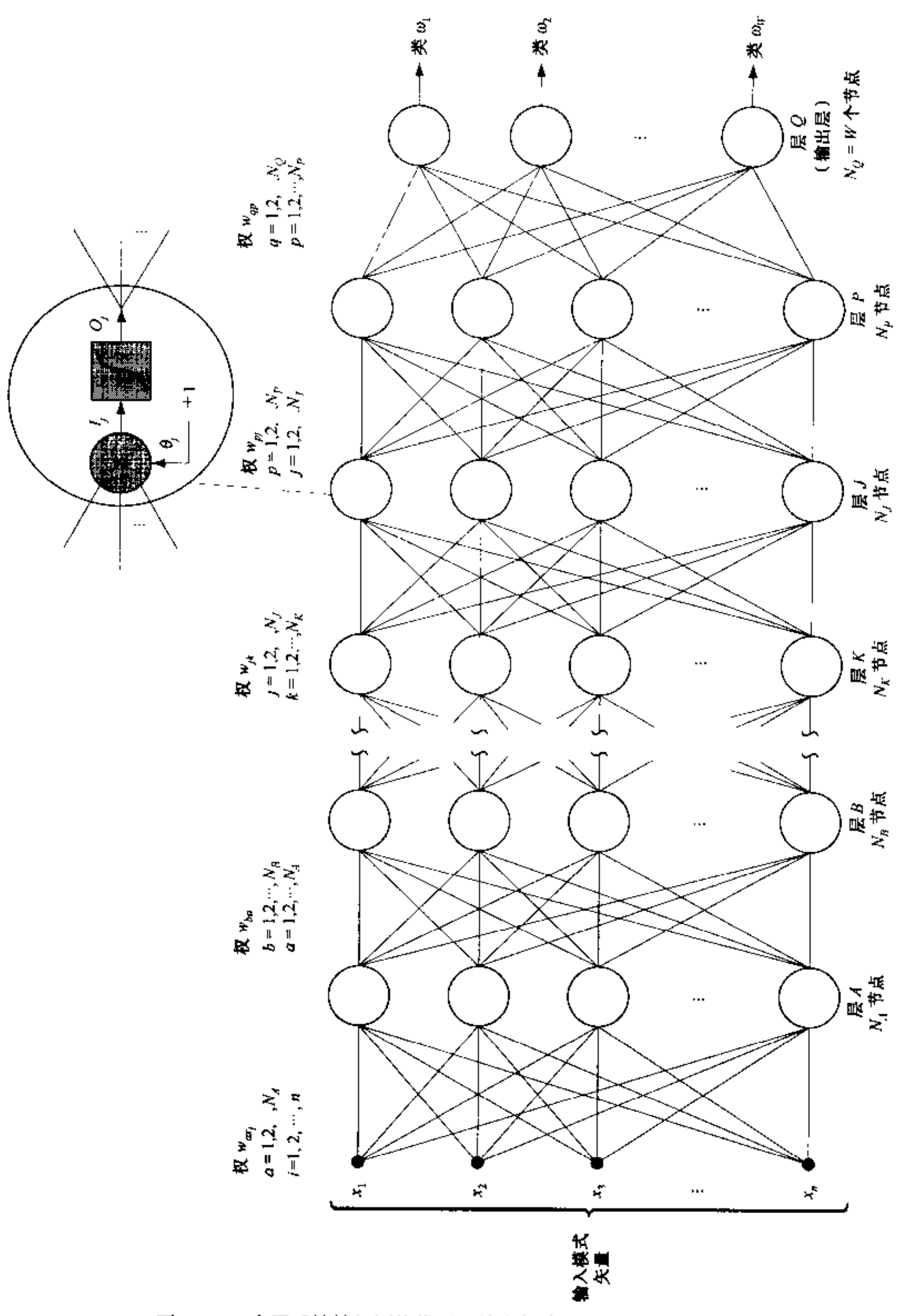


图 12.16 多层反馈神经网络模型。放大部分显示了整个网络中每个神经元的基本结构，偏移量θ_j当做不同的权值处理

这里 $I_j, j=1,2,\cdots,N_J$，是输入到网络第 J 层每一个节点的激活元素，θ_j 是偏移量，θ_o 控制“S”函数的形状。

图12.17中沿着每个节点响应的“高”和“低”输出描绘了式(12.2.47)。因此，当使用这个特殊函数时，系统对所有大于 θ_j 的 I_j 值输出一个“高”读数。同样，系统对所有小于 θ_j 的 I_j 值输出一个“低”读数。如图12.17所示，“S”激活函数总是正的，而且只有在激活元素分别为负无穷或正无穷时，才趋近于它的极限值0和1。由于这个原因，在图12.16中，用0和1的近似值(比如0.05和0.95)定义神经元输出的“低”和“高”值。原则上，不同类型的激活函数用于不同的层次，甚至用于神经网络同一层中的不同节点。实际上，通常的方法是对整个网络使用同样形式的激活函数。

参考图12.14(a)，图12.17中显示的偏移量与以前讨论感知器时谈到的权值系数 w_n+1 类似。要实现这种门限函数的替换可以如图12.14(a)中所示的形式，吸收偏移量 θ_j 作为修正输入网络节点的单位常量的附加系数。为了遵循在文献中普遍使用的符号，不把 +1 这个单独的常量输入显示在图12.16的所有节点中。代之而来的，这个输入及它的修正权值 θ_j 是网络节点的整数对。正如在图12.16中放大显示的，J 层中的 N_J 个节点的每一个都有一个这样的系数。

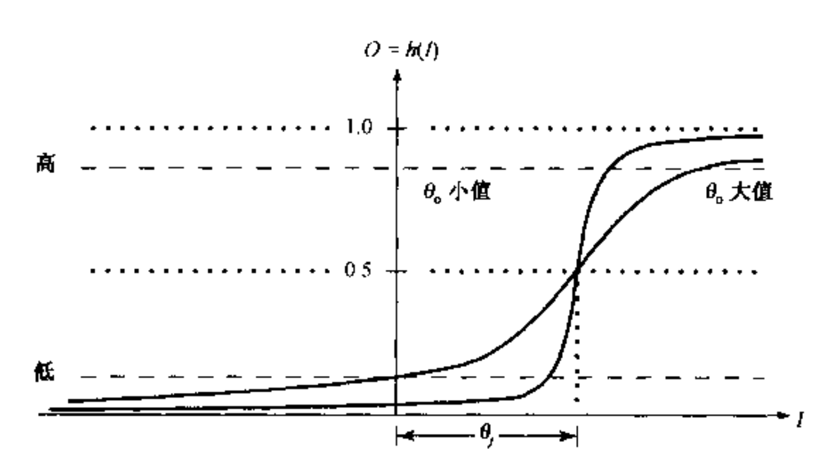


图12.17　式(12.2.47)中的激活函数“S”

在图12.16中，任何层中的节点输入都是前一层输出的加权和。令层 K 代表层 J 的前一层(在图12.16中，层的划分并不按字母顺序)。由层 K 生成层 J 的每一个节点的激活元素，用 I_j 代表：

$$I_j = \sum_{k=1}^{N_k} w_{jk} O_k \tag{12.2.48}$$

$j=1,2,\cdots,N_J$，这里 N_J 是层 J 中节点的数目，N_K 是层 K 中节点的数目，w_{jk} 是层 K 中节点的输出 O_k 馈给层 J 中节点之前，在层 K 中修正节点输出 O_k 的权值。层 K 的输出为：

$$O_k = h_k(I_k) \tag{12.2.49}$$

$k=1,2,\cdots,N_k$。

对式(12.2.48)所用的下标符号有一个清晰的理解是很重要的。因为在这一节的余下部

分都将用到它。首先,注意 $I_j, j = 1,2,\cdots,N_J$,表示层 J 第 j 个节点的激活元素输入。因此,I_1 表示层 J 第 1 个(顶端)节点的激活元素输入,I_2 表示层 J 第 2 个节点的激活元素输入,以此类推。对层 J 的每一个节点都有 N_K 个输入,但每个单个输入的加权都不同。因此,层 J 中第 1 个节点的输入 N_k 由系数 $w_{1k}(k = 1,2,\cdots,N_K)$进行加权;层 J 第 2 个节点的输入用系数 $w_{2k}(k = 1,2,\cdots,N_K)$进行加权;以此类推。所以,当层 K 的输出输入层 J 时,$N_J \times N_K$ 个系数作为层 K 输出的权值是必要的。一个附加的 N_J 偏移量系数 θ_j 需要用来完全确定层 J 中的节点。

将式(12.2.48)代入式(12.2.47)得到:

$$h_j(I_j) = \frac{1}{1 + e^{-\left(\sum_{k=1}^{N_K} w_{jk}O_k + \theta_j\right)/\theta_o}} \tag{12.2.50}$$

这是在本节余下部分中要用到的激活函数形式。

训练过程中,在输出层中修改神经元是很简单的,因为希望得到的每个节点的输出是已知的。训练一个多层网络过程中的主要问题在于,要调整称为隐藏层的权值。即,除了输出层以外的那些层中的权值。

反向传播的训练

我们先来关注输出层。输出层 Q 各节点的期望响应 r_q 和真实响应 O_q 之间总误差的平方为:

$$E_Q = \frac{1}{2}\sum_{q=1}^{N_Q}(r_q - O_q)^2 \tag{12.2.51}$$

这里 N_Q 是输出层 Q 的节点数目,系数$\frac{1}{2}$使后面求导时在表示上比较便利。

目的是推导类似于德尔塔规则的训练规则。使用这种训练规则可以调整每一层的权值,以便找到式(12.2.51)所示形式的误差函数最小值。像前面一样,与误差相对于权值的偏导数成比例地调整权值可以达到这个目的。换句话说,

$$\Delta w_{qp} = -\alpha\frac{\partial E_Q}{\partial w_{qp}} \tag{12.2.52}$$

这里层 P 先于层 Q,Δw_{qp}如式(12.2.42)所定义的,α 是正的校正增量。

误差 E_Q 是输出 O_q 的函数,O_q 则是输入 I_q 的函数。使用连锁规则,求出 E_Q 的偏导数,如下所示:

$$\frac{\partial E_Q}{\partial w_{qp}} = \frac{\partial E_Q}{\partial I_q}\frac{\partial I_q}{\partial w_{qp}} \tag{12.2.53}$$

由式 (12.2.48),得到:

$$\frac{\partial I_q}{\partial w_{qp}} = \frac{\partial}{\partial w_{qp}}\sum_{p=1}^{N_p} w_{qp}O_p = O_p \tag{12.2.54}$$

将式(12.2.53)和式(12.2.54)代入式(12.2.52)得到:

$$\Delta w_{qp} = -\alpha \frac{\partial E_Q}{\partial I_q} O_p$$
$$= \alpha \delta_q O_p \quad (12.2.55)$$

这里

$$\delta_q = -\frac{\partial E_Q}{\partial I_q} \quad (12.2.56)$$

为了计算偏导数$\partial E_Q / \partial I_q$,用连锁规则使用 E_Q 相对于 O_q 的变化率和 O_q 相对于 I_q 的变化率表示偏导数:

$$\delta_q = -\frac{\partial E_Q}{\partial I_q} = -\frac{\partial E_Q}{\partial O_q}\frac{\partial O_q}{\partial I_q} \quad (12.2.57)$$

由式 (12.2.51)得到:

$$\frac{\partial E_Q}{\partial O_q} = -(r_q - O_q) \quad (12.2.58)$$

从式 (12.2.49)得到:

$$\frac{\partial O_Q}{\partial I_q} = \frac{\partial}{\partial I_q} h_q(I_q) = h_q'(I_q) \quad (12.2.59)$$

将式(12.2.58)和式(12.2.59)代入式(12.2.57)给出:

$$\delta_q = (r_q - O_q) h_q'(I_q) \quad (12.2.60)$$

它与误差量值($r_q - O_q$)成比例。将式(12.2.56)到式(12.2.58)代入式(12.2.55)最终得到:

$$\Delta w_{qp} = \alpha(r_q - O_q) h_q'(I_q) O_p = \alpha \delta_q O_p \quad (12.2.61)$$

确定了函数 $h_q(I_q)$后,式(12.2.61)中的所有项都是已知的或者可以在网络中见到。换句话说,对于网络输入的任何训练模式的表达,我们知道每一个输出节点 r_q 的期望响应应该是什么。每个输出节点的值 O_q 可以被看做I_q(输入到层 Q 的激活元素)和 O_p(层 P 的节点输出)。因此,知道如何调整修改网络最末一层和次末层间链接的权值。

继续回到输出层的处理上来,现在分析层 P 发生的情况。以上述同样的方式进行处理,产生:

$$\Delta w_{pj} = \alpha(r_p - O_p) h_p'(I_p) O_j = \alpha \delta_p O_j \quad (12.2.62)$$

误差项是:

$$\delta_p = (r_p - O_p) h_p'(I_p) \quad (12.2.63)$$

除了 r_p,式(12.2.62)和式(12.2.63)中的所有项都已知或可在网络中见到。项 r_p 在内层是无意义的,因为不知道模式成员项中的内部节点响应应该是什么。可以仅在网络最终产生模式分类输出的地方指定所希望的响应项 r。如果知道内部节点的信息,就不需要下一层的信息。因此,必须找到一种方法在量上重定义 δ_p,使其成为已知的或在网络中可见的。

回到式(12.2.57),对层 P 写出误差项为:

$$\delta_p = -\frac{\partial E_p}{\partial I_p} = -\frac{\partial E_p}{\partial O_p}\frac{\partial O_p}{\partial I_p} \tag{12.2.64}$$

项$\partial O_p/\partial I_p$不难表示。以前,它表示成:

$$\frac{\partial O_p}{\partial I_p} = \frac{\partial h_p(I_p)}{\partial I_p} = h_p'(I_p) \tag{12.2.65}$$

我们知道,h_p 是已确定的,因为 I_p 是可观测的。产生 r_p 的项是偏导数$\partial E_p/\partial O_p$,所以这一项必须以不包含 r_p 的方法表示。使用连锁规则,将偏导数写成:

$$\begin{aligned} -\frac{\partial E_p}{\partial O_p} &= -\sum_{q=1}^{N_Q}\frac{\partial E_p}{\partial I_q}\frac{\partial I_q}{\partial O_q} = \sum_{q=1}^{N_Q}\left(-\frac{\partial E_p}{\partial I_q}\right)\frac{\partial}{\partial O_p}\sum_{p=1}^{N_p} w_{qp}O_p \\ &= \sum_{q=1}^{N_Q}\left(-\frac{\partial E_p}{\partial I_q}\right)w_{qp} \\ &= \sum_{q=1}^{N_Q}\delta_q w_{qp} \end{aligned} \tag{12.2.66}$$

最后一步来自式(12.2.56)。将式(12.2.65)和式(12.2.66)代入式(12.2.64)得到对 δ_p 所希望的表达式:

$$\delta_p = h_p'(I_p)\sum_{q=1}^{N_Q}\delta_q w_{qp} \tag{12.2.67}$$

因为所有项已知,现在参量 δ_p 是可计算的。因此,式(12.2.62)和式(12.2.67)完全完成了层 P 的训练法则。式(12.2.67)的重要性在于从 δ_q 和 w_{qp} 计算出 δ_p,这两个量是在紧接着层 P 的层中计算出来的项。对于层 P,误差项和权值计算出来后,这些量也可以类似地用于直接计算层 P 之前层的误差项和权值。换句话说,已经找到一种方法,从输出层误差开始把误差反向传播到网络。

我们可以总结和归纳训练过程如下。对任何层 K 和 J,层 K 直接接在层 J 的前边,计算权值 w_{jk},w_{jk}通过使用下列公式,改进了两层的连接:

$$\Delta w_{jk} = \alpha\delta_j O_k \tag{12.2.68}$$

如果层 J 是输出层,δ_j 为:

$$\delta_j = (r_j - O_j)h_j'(I_j) \tag{12.2.69}$$

如果层 J 是内部层,并且层 P 是它的下一层(右边),则 δ_j 由下式给出:

$$\delta_j = h_j'(I_j)\sum_{p=1}^{N_p}\delta_p w_{jp} \tag{12.2.70}$$

$j=1,2,\cdots,N_j$。使用式(12.2.50)中的激活函数,并令 $\theta_0=1$ 得到:

$$h_j'(I_j) = O_j(1-O_j) \tag{12.2.71}$$

在这种情况下,式(12.2.69)和式(12.2.70)遵循如下假定时,有特别吸引人的形式:

$$\delta_j = (r_j - O_j)O_j(1-O_j) \tag{12.2.72}$$

对应于输出层,有:

$$\delta_j = O_j(1 - O_j)\sum_{p=1}^{N_p} \delta_p w_{jp} \tag{12.2.73}$$

对应于内部层。在式(12.2.72)和式(12.2.73)中,$j = 1,2,\cdots,N_j$。

式(12.2.68)到式(12.2.70)构成图 12.16 的训练多层反馈神经网络的推广德尔塔法则。这个过程由一个通过整个网络的任意(但不全相等)的权重集合开始。然后,推广的德尔塔法则用于任何涉及两个基本阶段的迭代步骤。第一个阶段,对于网络存在一个训练矢量,并且容许层传播对每个节点计算其输出。输出层节点的输出 O_q 与对它们期望的响应值 r_p 进行比较,生成误差项 δ_q。第二阶段包括一条遍历网络的反向路径,在此阶段,合适的误差信号通过每个节点并且相应的权重做出改变。这个过程同样应用于权值偏离量 θ_j。在以前讨论的细节中,这些权值偏离量仅被看做附加权值,作为输入到网络中每个节点总连接的单位量修正值。

通常的做法是对网络误差和与单个模式相联系的误差进行跟踪。在成功的训练中,网络误差随迭代步骤次数的增加而减少并且这一过程收敛于一个稳定的权值集合,它对附加训练仅呈现很小的影响。接下来要完成的方法需要确定一个模式是否已被正确分类,训练时确定与获得模式的模式类有关的输出层节点响应是否是高的,如以前定义的那样,所有其他节点的输出是低的。

系统进行训练之后,使用在训练阶段中设定的参量对模式进行分类。在标准操作中,所有反馈路径是不连通的。任何输入模式允许通过不同层进行传播,并且模式被划归输出为高的节点所属的类。此时,其他所有节点输出为低。如果被标记为高的节点不止一个,或没有节点输出被标记为高,则可选的做法是,声明进行了错误的分类或简单地将模式划归输出节点的类并赋予最大值。

例 12.6 使用神经网络进行图形分类

现在说明如何训练一个如图 12.16 所示形式的神经网络去识别图 12.18(a)所示的 4 种形状及这些形状带有噪声的类型,样本如图 12.18(b)所示。

通过计算形状的归一化信号(见 11.1.3 节)得到每个信号的 48 个隔开的均匀分布样本来产生模式矢量。得到的 48 维矢量输入到图 12.19 所示的三层前馈神经网络。第一层神经元节点的数目定为 48,与输入模式矢量的维数一致。第三层(输出层)的四个节点与模式类的数目一致,并且中间层的神经元数目确定为 26(输入和输出层神经元的平均数)。由于不知道确定神经网络内部层次中节点数目的规则,因此这个数目一般是基于以前的经验或任意指定并通过检验来完善。在输出层,从上到下的四个节点此时各代表类 ω_j, $j = 1,2,3,4$。在设定网络结构后,必须为每个单元和层选定激活函数。所有选定的激活函数要满足当 $\theta_o = 1$ 时的式(12.2.50),根据以前的讨论,要使用式(12.2.72)和式(12.2.73)。

训练过程分为两个部分。在第一部分中,权值被初始化为带有零均值的小随机数,然后,使用与类似于图 12.18(a)所示的相应无噪声样本的模式矢量对网络进行训练。外形图如图 12.18(a)中所示。输出节点在训练期间是受到监控的。对类 w_i 的所有训练模式,当输

出层的元素产生输出值 $O_i \geqslant 0.95$ 和 $O_q \leqslant 0.05$ 时($q = 1,2,\cdots,N_Q$ 且 $q \neq i$),我们说网络从所有的4个类学到了这些形状。换句话说,对类 ω_i 的所有模式,与所求类一致的输出节点必须为高($\geqslant 0.95$),而同时,所有其他节点必须为低($\leqslant 0.05$)。

训练的第二部分要在带有噪声样本的情况下训练。在无噪声图形中,每个处在轮廓线上的像素被给予一个保留在图形平面上的原始坐标的概率 V,并且对该像素8个相邻像素其中一个的坐标随机赋予一个概率 $R = 1 - V$。噪声的等级随 V 的减小而增加(即增加 R)。两个噪声数据集合就产生了。第一个集合通过在0.1到0.6之间的不同 R 值对每个类生成100个噪声模式,总共为400个模式。这个集合称为检测集合,用于确定系统在训练后的性能。

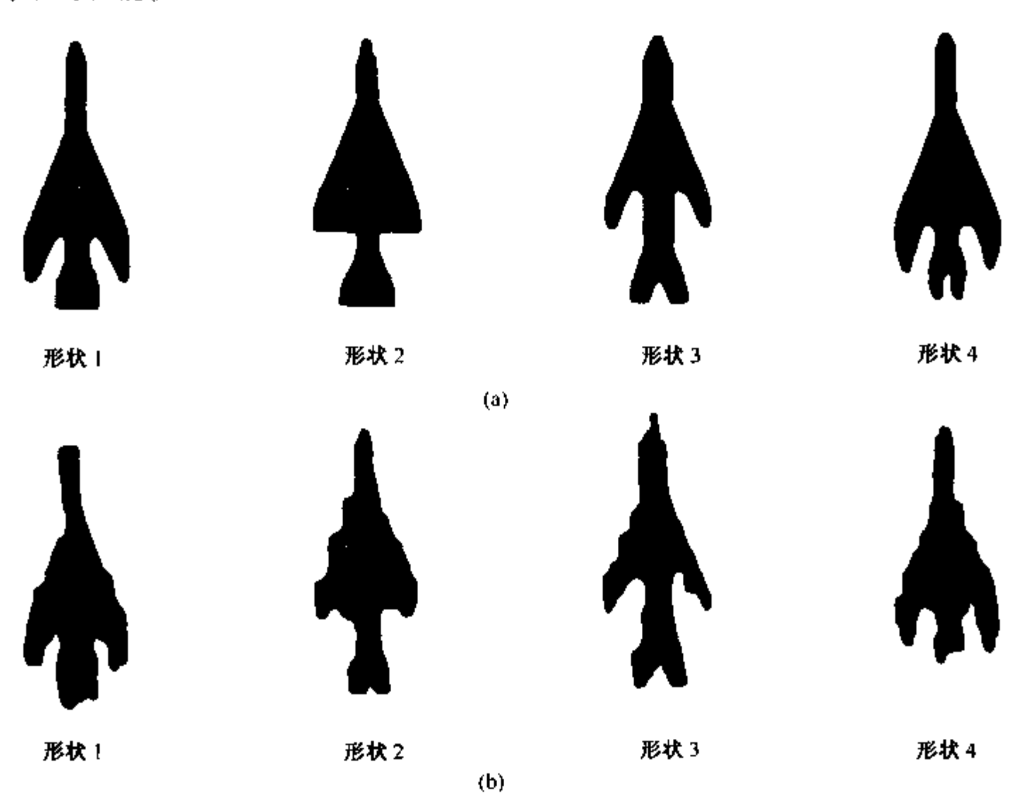


图12.18 (a)参考图形和(b)用于训练图12.19所示神经网络的典型噪声图形(由南伊利诺伊大学ECE系的Lalit Gupta博士提供)

用噪声数据训练系统要产生几个噪声集合。第一个集合由每类的10个样本组成,通过令 $R_t = 0$ 生成,这里 R_t 代表用于生成训练数据的 R 值。从第一部分(无噪声)训练得到的权值矢量开始,系统允许经过一系列有序的带有新数据集合的学习过程。因为 $R_t = 0$ 暗示无噪声,这种再训练是以前无噪声训练的延伸。以这种方式产生学习过的权重,网络受测试数据集的影响,数据集生成的结果如图12.20中标有 $R_t = 0$ 的曲线所示。错误分类的模式数目除以检测过的模式总数得到错误分类的概率,一般是衡量确立的神经网络性能的度量。

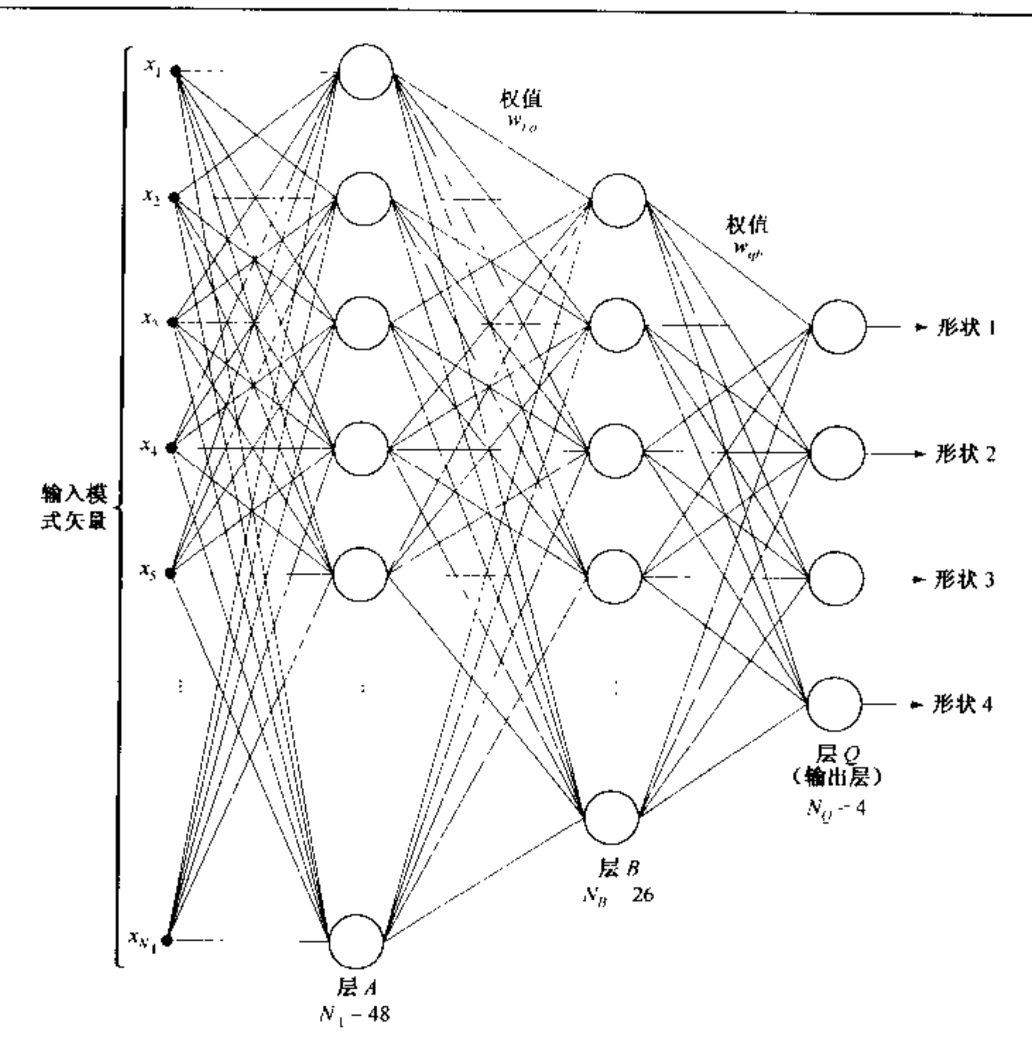


图 12.19 用于识别图 12.18 中图形的三层神经网络(由南伊利诺伊大学ECE系的Lalit Gupta博士提供)

接下来,从使用 $R_t=0$ 生成的噪声数据权值矢量开始,用由 $R_t=0.1$ 生成的噪声数据集合对系统进行重新训练。识别性能再用新权值矢量通过系统运行检测样本来确立。注意,这次系统在性能上的显著改进。图 12.20 显示了通过持续这种再训练和令 $R_t=0.2,0.3$ 和 0.4 后进行的再检测过程得到的结果。如所希望的那样,如果系统进行了适当的学习,由检测集合测试出的模式错误分类的概率会在 R_t 增加时降低,因为系统使用了有更高 R_t 值的噪声数据进行训练。图 12.20 中的一个例外是 $R_t=0.4$ 时的结果。原因是训练系统的样本数目少,即网络无法在使用这样少的样本情况下使自己充分适应具有更高噪声级别的图形的更大变化。这种假设在图 12.21 中得到证实,在此图中显示出,当样本数目增加时得到更低的错误分类概率。图 12.21 也显示了图 12.20 中的曲线在 $R_t=0.3$ 时的形状作为参考。

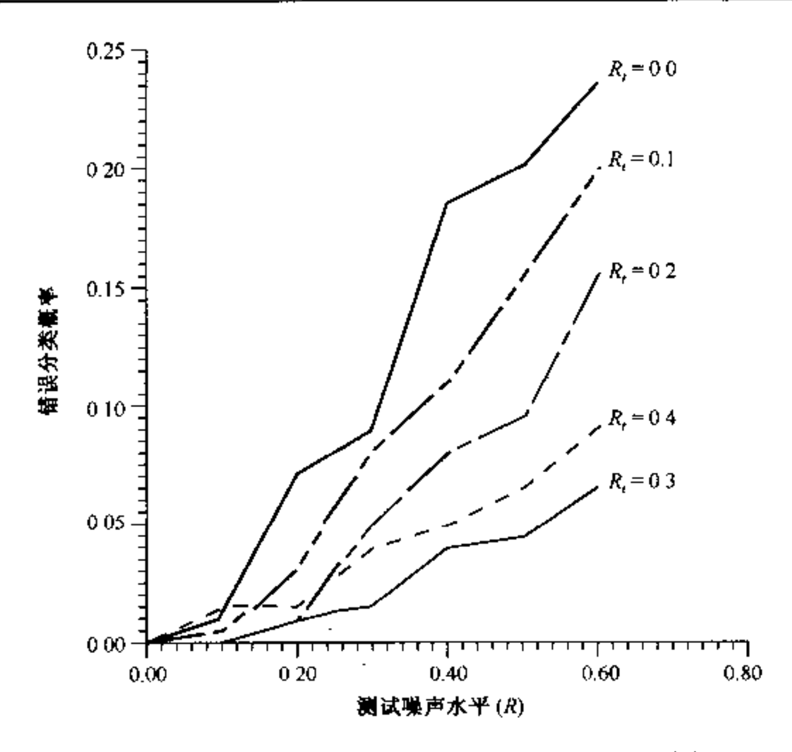


图 12.20 作为噪声水平函数的神经网络性能（由南伊利诺伊大学ECE系的Lalit Gupta博士提供）

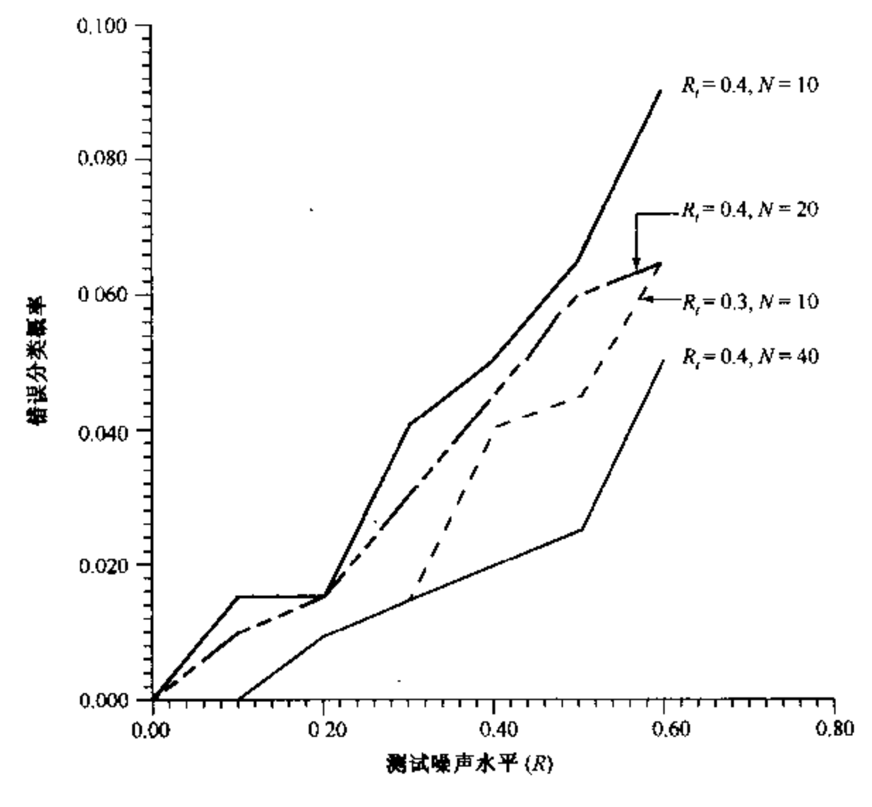


图 12.21 通过增加训练模式数目改进 $R_t = 0.4$ 时的效果($R_t = 0.3$ 的曲线作为参考)(由南伊利诺伊大学ECE系的Lalit Gupta博士提供)

前述结果显示出一个三层神经网络在经过适当训练后有能力识别被噪声干扰的图形。甚至使用无噪声数据(图 12.20 中 $R_t=0$)训练的系统当使用被较高噪声(图 12.20 中 $R=0.6$)干扰的数据测试时,其正确识别的水平依然可以接近 77%。当系统使用噪声水平更高的数据($R_t=0.3$ 和 $R_t=0.4$)训练后,对同样数据的识别率增加到 99%。应该注意到通过系统性的、噪声水平小量增加的训练来增强系统的分类能力是很重要的。当噪声性质已知时,这种方法对改进神经网络在学习过程中的收敛和稳定性是很理想的。

决策面的复杂性

我们已经确定一个单层感知器执行一个超平面的决策面。关于这一点一个自然而然的问题是,像图 12.16 中的模型那样,由一个多层网络实现的决策面性质是什么? 这个问题将在下面的讨论中论证,下面的讨论是一个三层网络有能力实现由相交超平面组成的任意复杂的决策面。

作为讨论的起点,考虑图 12.22(a)所示的两个输入并有两层的网络。具有两个输入,模式是两维的,所以网络第一层的每个节点执行的是二维空间中的一行。用 1 和 0 分别代表这两个节点的高和低输出。假设 1 输出表示,对第一层一个节点的相应输入向量位于直线为正的一边。然后,有可能输出给第二层中单个节点的组合是(1,1),(1,0),(0,1)和(0,0)。如果定义两个区域,一个对应类 ω_1,位于两条直线正的一侧,而另一个对应于类 ω_2,位于直线的另一侧,输出节点可以把任何输入模式分类,只需要对属于这两个区域之一的输入模式进行简单的逻辑"与"操作。换句话说,输出节点响应为 1,仅当第一层的两个输出都为 1 时,表示类 ω_1。如果 θ_j 在半开半闭区间(1,2]中设值,"与"操作可以由早些时候讨论的形式的神经节点来执行。如果假设为 0 和为 1 的响应超出第一层,只有当来自第一层的神经节点的两个输出执行的求和大于 1 时,输出节点的响应才会为高,表示类 ω_1。图 12.22(b)和(c)显示了图 12.22(a)所示网络如何成功地等分两个模式类,而使用一个单一的线性面则做不到这一点。

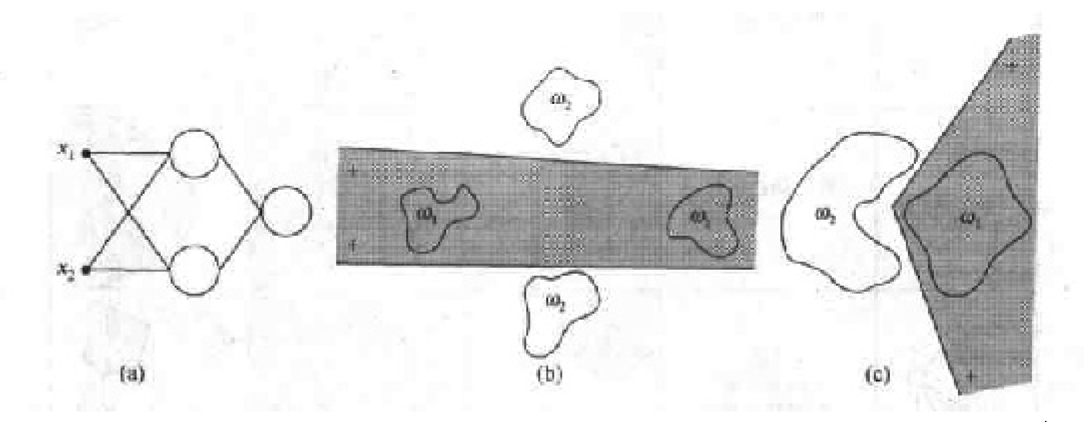


图 12.22　(a)一个两输入、两层的前馈神经网络。(b)和(c)可以由这个网络实现的决策边界的例子

如果第一层中的节点数目增加到 3 个,图 12.22(a)中的网络将由 3 条直线相交组成的判别边界来执行。类 ω_1 位于所有 3 条直线正的一侧的要求产生一个由这 3 条直线作为边界线的凸区域。实际上,任何开放的或闭合的凸区域都可以简单地通过增加两层神经网络第一层的节点数来构造。

下一个逻辑步骤是将层数增加到3层。在这种情况下,和以前一样,第一层的节点执行一行。为了从各个行形成一个区域,第二层的节点进行“与”操作。第三层的节点将不同区域划归不同的类。例如,设想类 ω_1 由两个不同区域组成,每个区域由不同的直线集合构成边界。第二层节点中的两个对应于相同模式类的区域。当第二层的两个节点中任一个变为高时,输出节点中应有一个能以信号表示该类的存在。假设第二层中由1和0分别代表高和低状态,这可以通过令网络的输出节点执行逻辑“或”操作得到。根据前面讨论过的神经节点的形式,给 θ_j 在半开区间内设置一个值。无论什么时候,在第二层中至少有一个节点与变为高状态的输出节点(输出为1)相联系,输出层的相应节点变为高状态就表示被处理的模式属于与那个节点相联系的类。

图12.23总结了前面的讨论。注意在第三行中,由三层网络实现的判别区域的复杂性在原理上是任意的。实际上,通常一系列的困难主要是由构造第二层来正确响应与各种特殊类相关的组合产生的。原因是直线不会正好在与其他直线的交点处截止,结果在模式空间中,属于相同类的模式也许会在直线的两边出现。从实际方面讲,在第二层中很难画出哪一条线应该包括在给定模式类的“与”操作中,或许这根本就是不可能的。参考图12.23的第三列中异或问题处理的情况,即,当输入模式为二元形式时,以两个维度只能构造四个不同的模式。如果模式被排列成如下形式:类 ω_1 由模式{(0,1),(1,0)}构成,类 ω_2 由模式{(0,0),(1,1)}构成,这两个类中,模式的类归属由异或(XOR)逻辑函数给出。这个函数只有在两个变量中的一个为1时才为1,否则就为0。因此,XOR函数值为1表示模式属于类 ω_1,XOR函数值为0表示模式属于类 ω_2。

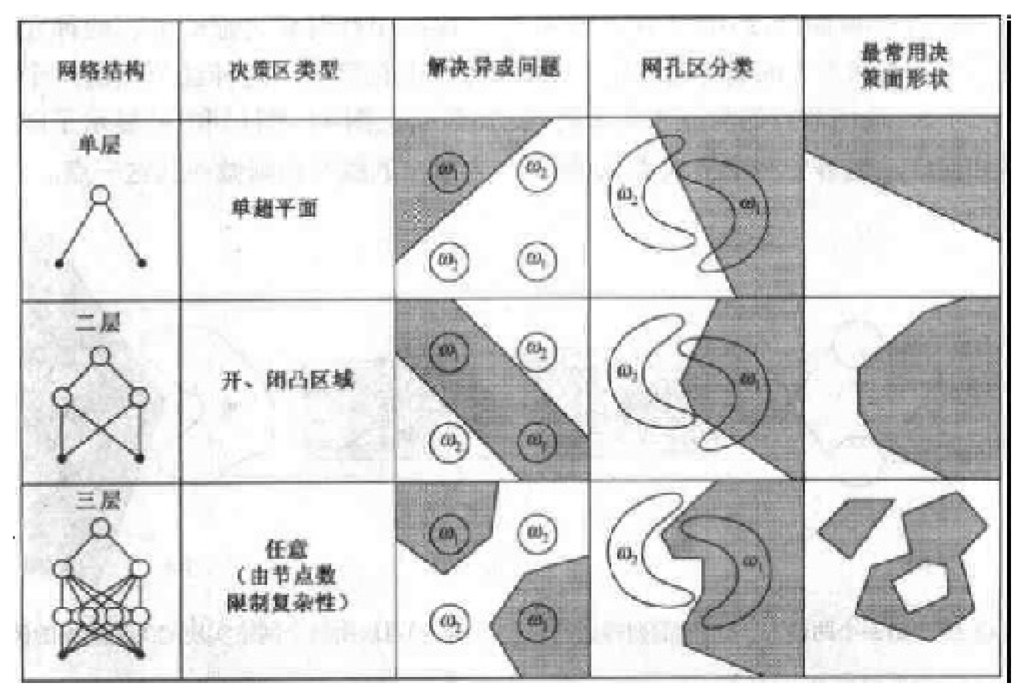


图12.23 可以由单层和多层前馈网络组成的判别区域类型,这个网络有一层或两层隐藏单元和两个输入(Lippman)

前面的讨论以直接的方式推广到 n 维情况:这种情况中处理的是超平面而不是直线。一个单层网络执行一个单一的超平面。一个两层网络执行由超平面交集组成的任意凸区域。一

个三层网络执行任意复杂的判别面。每层使用的节点数决定了后两种情况的复杂性。第一种情况中，限制类的数目为两个。在另两种情况下，类的数目是任意的，因为可以根据要处理的问题决定输出节点的数目。

考虑到前述的评论，一个合乎逻辑的问题是，为什么人们对研究三层以上的神经网络有兴趣？毕竟，一个三层网络可以执行任意复杂的决策面。答案就在于仅使用三个层次对网络进行训练时所使用的方法上。图 12.16 中网络的训练规则将误差度量降至了最小，但却并未谈到如何把多个超平面和以前讨论过的三层网络第二层中特殊的节点联系起来。实际上，如何在层次数和每层上的节点数之间进行折中的分析的问题仍没有解决。在实现过程中，一般是通过反复实验和检验错误或凭借对给定问题领域的经验解决折中分析难题的。

12.3 结构性方法

在 12.2 节里讨论过在数量上处理模式的技术。这种技术在很大程度上忽略了模式图形的任何结构上的关系。在这一节中，将讨论结构性方法，通过对这些类型的关系进行适当的估计来实现模式识别。

12.3.1 匹配形状数目

在 12.2.1 节中，为了将模式矢量公式化，介绍了最小距离的概念。模式矢量公式化是为了对比依据图形数目描绘的区域边界。以 11.2.2 节的论述为参考，两个区域边界间(形状)的相似度 k 定义为它们相一致的图形数目的最大量级。例如，令 a 和 b 代表有闭合边界的图形数目。闭合边界通过 4 方向的链码表示。如果：

$$\begin{aligned} s_j(a) &= s_j(b) \qquad j = 4,6,8,\cdots,k \\ s_j(a) &\neq s_j(b) \qquad j = k+2,k+4,\cdots \end{aligned} \tag{12.3.1}$$

则这两个图形有一个相似程度 k。这里 s 代表图形数，下标代表次序。两个图形 a 和 b 间的距离定义为它们相似程度的倒数：

$$D(a,b) = \frac{1}{k} \tag{12.3.2}$$

a 和 b 间的距离满足下列性质：

$$\begin{aligned} &D(a,b) \geqslant 0 \\ &D(a,b) = 0，\text{当且仅当 } a = b \\ &D(a,c) \leqslant \max[D(a,b),D(b,c)] \end{aligned} \tag{12.3.3}$$

无论 k 或 D 都可以用于两个图形的比较。如果应用了相似度，则 k 越大，图形越相似(注意 k 对同样的图形是无穷大的)。当使用距离度量时与使用 k 时相反。

例 12.7 应用图形数目对比图形

假设有一个图形 f，并且希望在有 5 个其他图形(a,b,c,d 和 e)的集合中找到与它最为接近的匹配，如图 12.24(a)所示。这个问题同已有 5 个原型图形而试图寻找给定未知图形的最佳匹配相似。有图 12.24(b)中显示的相似树的辅助，这个搜索有可能被可视化。树根节点对应相似度最低的可能性，在这个例子中是 4。假设图形相似程度为 8，除了图形

a(它与其他所有图形的相似程度为 6)。沿着树向下,发现图形 d 与其他图形的相似程度为 8,等等。

图形 f 和 c 惟一地匹配,有高于其他任何两个图形的相似度。在其他极端情况下,如果 a 是一个未知图形,那么可以说,使用这种方法使 a 同其他 5 个图形以相似度为 6 的程度相似。使用相似性矩阵可以得到同样的结论。相似性矩阵如图 12.24(c)所示。

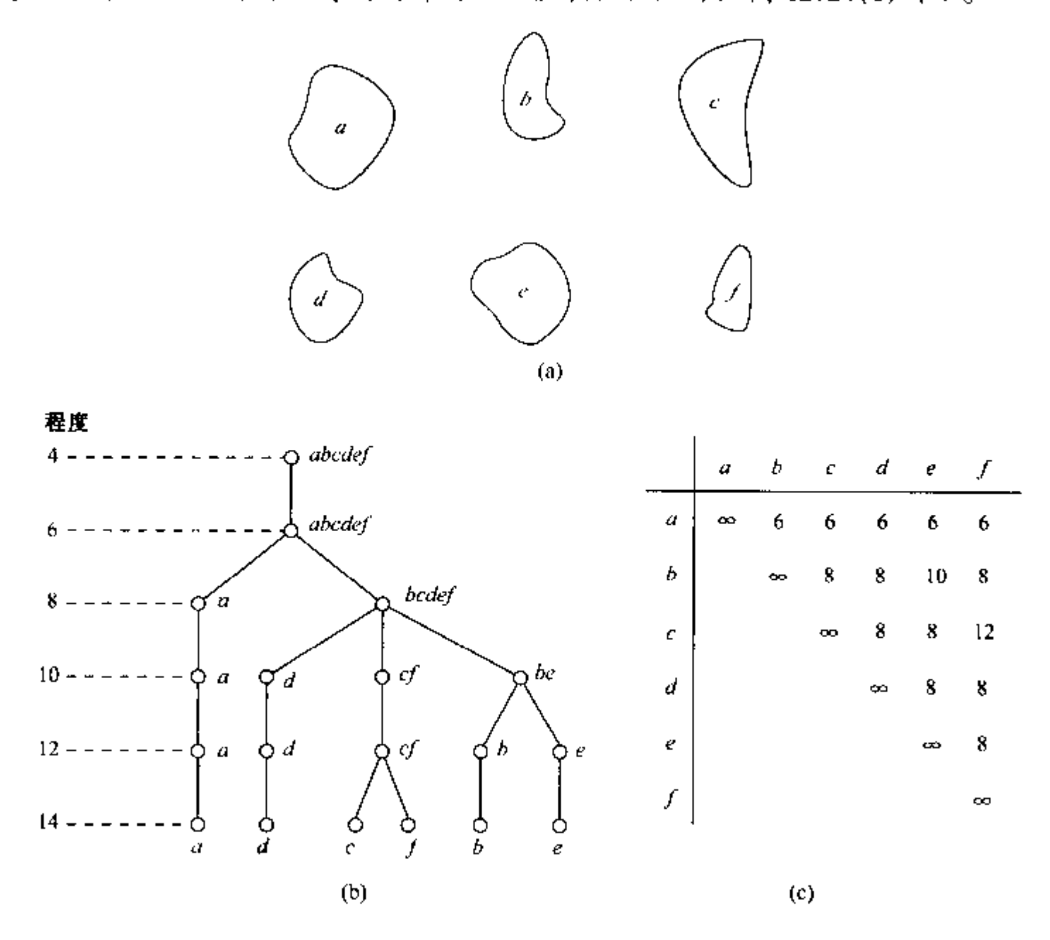


图 12.24 (a)图形,(b)假定的相似树,(c)相似矩阵(Bribiesca 和 Guzman)

12.3.2 串匹配

假设有两个区域边界 a 和 b,被编码后写成串(见 11.5 节),分别用 $a_1a_2,\cdots,a_n$ 和 $b_1b_2,\cdots,b_m$ 表示。令 α 代表两个串之间的匹配数目,如果 $a_k=b_k$,则匹配发生在第 k 个位置上。无法匹配的字符数目为:

$$\beta = \max(|a|,|b|) - \alpha \tag{12.3.4}$$

这里|arg|是串码变量的长度(字符数目)。当且仅当 a 和 b 相同时 $\beta=0$(见习题 12.21)。

一种判断 a 和 b 间相似性的简单方法是使用如下所示的比率:

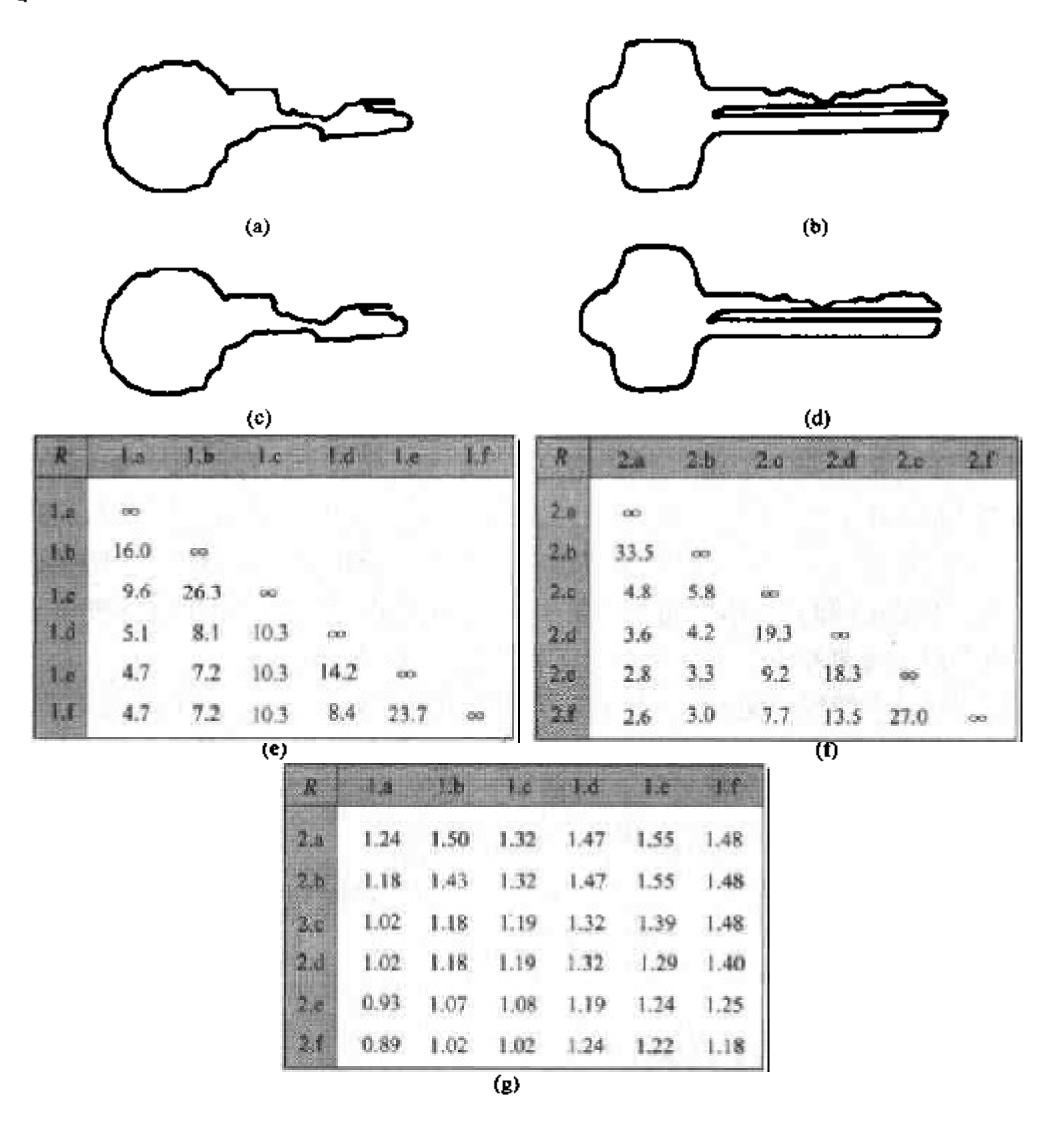
$$R=\frac{\alpha}{\beta}=\frac{\alpha}{\max(|a|,|b|)-\alpha} \tag{12.3.5}$$

对完全匹配 R 为无限大,a 和 b 没有任何字符匹配时 R 为 0(此时 $\alpha=0$)。因为匹配是逐个字符

进行的，对于减少计算量来说，每条边界的起始点是很重要的。对任何标准化的或近似标准化的方法，同样的起点是很有帮助的，只要这种方法比强制匹配[从包括在每一串的任意点开始，移动串的一点，并对每一次移动计算式(12.3.5)]有计算上的优点。R 的最大值给出最佳匹配。

例 12.8　串匹配的说明

图 12.25(a)和(b)显示了来自两个对象类的样本边界，它们使用多边形拟合近似给出(见 11.1.2 节)。图 12.25(c)和(d)分别显示了对应于图 12.25(a)和(b)中边界的多边形近似。串是由计算每一个多边形被顺时针追踪时各段间的内角形成的。将角度编成用 8 个字符之一代表的码，对应于 45°的增量，即 $\alpha_1: 0° < \theta \leqslant 45°; \alpha_2: 45° < \theta \leqslant 90°; \cdots; \alpha_8: 315° < \theta \leqslant 360°$。



(e)

R	1.a	1.b	1.c	1.d	1.e	1.f
1.a	∞					
1.b	16.0	∞				
1.c	9.6	26.3	∞			
1.d	5.1	8.1	10.3	∞		
1.e	4.7	7.2	10.3	14.2	∞	
1.f	4.7	7.2	10.3	8.4	23.7	∞

(f)

R	2.a	2.b	2.c	2.d	2.e	2.f
2.a	∞					
2.b	33.5	∞				
2.c	4.8	5.8	∞			
2.d	3.6	4.2	19.3	∞		
2.e	2.8	3.3	9.2	18.3	∞	
2.f	2.6	3.0	7.7	13.5	27.0	∞

(g)

R	1.a	1.b	1.c	1.d	1.e	1.f
2.a	1.24	1.50	1.32	1.47	1.55	1.48
2.b	1.18	1.43	1.32	1.47	1.55	1.48
2.c	1.02	1.18	1.19	1.32	1.39	1.48
2.d	1.02	1.18	1.19	1.32	1.29	1.40
2.e	0.93	1.07	1.08	1.19	1.24	1.25
2.f	0.89	1.02	1.02	1.24	1.22	1.18

图 12.25　(a)和(b)是两个不同对象类的边界，(c)和(d)是它们对应的多边形近似，(e)到(g)是 R 值表(Sze 和 Yang)

图 12.25(e)显示了对象 1 的 5 个样本的度量值 R 与它们自身之比的值。例如，相应于 R 的全部值，符号 1.c 来自类 1 的第 3 个串。图 12.25(f)显示的结果是第 2 个对象类的串和

它们自身之比。最后,图 12.25(g)显示的是通过比较一个类和其他类的串得到的 R 值表。这里请注意,所有的 R 值比前两个表中的任何项都小。这表示 R 度量方法在两类对象间达到了很高的辨别级别。例如,如果串 1.a 的类别成员未知,将本串和类 1 的样本(原型)串相比较产生的最小 R 值将是 4.7[图 12.25(e)]。相比之下,将本串和类 2 的串相比较的最大值将会是 1.24[图 12.25(g)]。从这个结果可以得出结论:串 1.a 是对象类 1 的成员。这种分类方法和 12.2.1 节介绍的最小距离分类器相似。

12.3.3 串的语法识别

语法方法为处理结构识别问题提供了一种统一的方法。基本上,语法模式识别的主要思想是规定一个模式基元集合(见 11.5 节)、一个规定模式基元相互间关联的规则集合(采用语法的形式)和一个识别器(称为自动机),识别器的结构由语法规则集合决定。首先考虑串文法及其自动机,在下一节中将把这种思想扩展到树文法及其相应的自动机。

串文法

假设有两个类 ω_1 和 ω_2,它们的模式是 11.5 节所讨论方法生成的元素组成的串。可以把每个元素作为某种文法的字母表中可接受的字符,这里文法是一个句法规则的集合(因此有语法识别的名称)。这个句法集合规范着由字符生成的句子。文法 G 生成的句子集合叫语言,用 $L(G)$表示。这里,句子是符号组成的串(依次代表模式),语言与模式类相对应。

考虑两个文法 G_1 和 G_2,它们的句法规则是:G_1 只允许生成与来自类 ω_1 的模式相对应的句子,G_2 只允许生成与来自类 ω_2 的模式相对应的句子。带有上述特性的文法一经设定,句法模式识别过程在原理上就简单了。对于一个表示未知模式的句子,目的是判别在哪种语言下模式表示一个有效的句子。如果句子属于语言 $L(G_1)$,认为模式属于类 ω_1。同样,如果句子在语言 $L(G_2)$中是正确的,则认为此句子属于类 ω_2。如果句子同时属于这两种语言,就无法得到惟一的判别。对两种语言来说都不能正确识别的句子将被拒绝。

在存在两种以上的模式类时,句法分类方法与前面段落中描述的一样,只是过程中涉及的文法更多而已(每个类至少有一条对应的文法)。对多个类别的分类,如果它仅对于语言 $L(G_i)$是可用的句子,则模式属于类 ω_i。与前面一样,如果一个句子属于不止一种语言,就不能做出惟一的判别。所有语言都不可用的句子会被拒绝。

在处理串时,用四元组定义一个文法:

$$G = (N, \Sigma, P, S) \tag{12.3.6}$$

这里

N 是有限变元集合,称为非终端符,

Σ是有限常元集合,称为终端符,

P 是重写规则集合,称为产生式,

S 属于 N 集合,称为起始符。

这里要求 N 和Σ是不相交的集合。在下面的论述中,大写字母 $A, B, \cdots, S, \cdots$代表非终端符号。字母表中开头的小写字母 $a, b, c, \cdots$代表终端符号。字母表中末尾的小写字母 v, w, x, y, z 代表终端符号串。小写希腊字母 $\alpha, \beta, \theta, \cdots$代表终端符号和非终端符号混合组成的串。

空语句(没有符号的句子)用 λ 表示。最后,对于一个符号集合 V,符号 V^* 表示由来自 V 的元素组成的句子集合。

串文法由它的产生式描述其性质。在句法模式识别中,特别感兴趣的是正则文法和上下文无关文法。正则文法的产生式只有 $A \to aB$ 或 $A \to a$ 的形式。A 和 B 属于 N,a 属于 Σ。上下文无关文法的产生式形式为 $A \to \alpha$,A 属于 N、α 属于 $(N \cup \Sigma)^*$,即 α 是除空字符以外的终端符号和非终端符号构成的串。

例 12.9 使用正则串文法生成对象类

在开始之前,考虑一下文法生成对象类的机理是有用的。假设图 12.26(a)中显示的对象由它的(已删除的)骨架来描述,并且定义示于图 12.26(b)中的基元来描述这一(和相似的)骨架的结构。考虑文法 $G = (N, \Sigma, P, S)$,$N = \{A, B, S\}$,$\Sigma = \{a, b, c\}$ 且 $P = \{S \to aA, A \to bA, A \to bB, B \to c\}$,这里终端符号 a, b, c 对应于图 12.26(b)中所示的基元。如前边所指出的那样,S 是生成语言 $L(G)$ 的字串的起始符。例如,使用第一个产生式,并接下来使用两次第二个产生式得到:$S \Rightarrow aA \Rightarrow abA \Rightarrow abbA$,这里($\Rightarrow$)表示一个串从 S 开始推导,使用 P 集合中的产生式得到串。第一个产生式将 S 重写为 aA,第二个产生式将 A 重写为 bA。串 $abbA$ 中有一个非终端符,可以继续推导。例如,再使用两次第二个产生式,使用一次第三个产生式,并使用一次第四个产生式,生成的串 $abbbbbc$,对应图 12.26(c)所示的结构。使用第四个产生式后,串中就不存在非终端符了。因此,使用第四个产生式后推导就结束了。由此文法规则产生的语言是 $L(G) = \{ab^n c \mid n \geq 1\}$,这里 b^n 表示符号 b 重复 n 次。换句话说,G 仅能产生形如图 12.26(c)所示的骨架形式,但长度不限。

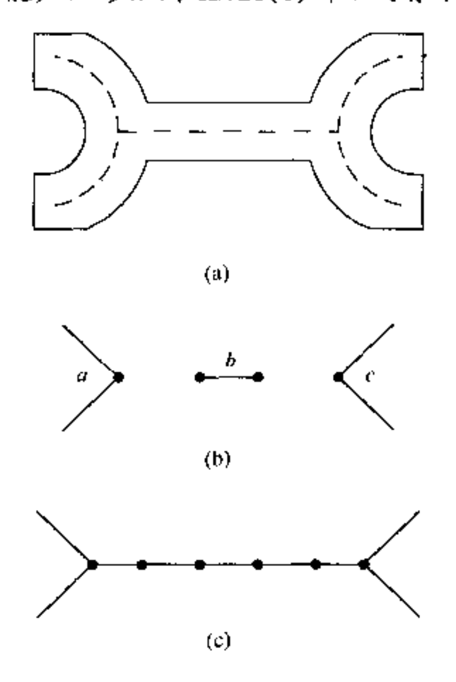


图 12.26 (a)由骨架代表的对象,(b)原始基元,(c)使用正则串文法生成的结构

语义学的使用

在前面的例子中,假定基元间的相互连接只在该点发生,如图 12.26(b)所示。在更为复杂的情况下,必须搞清楚连通性的规则、其他因素的有关信息(如基元的长度和方向)、产生式可以应用的次数等因素。这些可以通过存储在知识库(见图 1.23)中的语义规则来实现。基本上,在产生式规则中句法内在的规则设定了对象的结构,但语义学处理对象的正确性。例如,像 C 语言这样的编程语言,$A = D/E$ 在句法上是正确的,但只有在 $E \neq 0$ 时,在语义上才是正确的。

设想把语义信息和前面例子中谈到的文法结合起来。语义信息可以同表 12.1 中的产生式联系起来。通过使用语义信息,可以使用少量的句法规则描述广泛的类模式(但受我们的要求限制)。例如,通过指定表 12.1 中 θ 的方向,避免了在每个方向上指定元素。同样,通过要求所有元素定位于同一个方向,可以不去考虑由图 12.26(a)代表的偏离基本形状的无意义结构。

表 12.1 产生规则附带的语义信息实例

产生式	语义信息
$S \to aA$	与 a 的连接只在该点进行。表示为 θ 的 a 的方向由直线的垂直等分线给出。此直线连接两条没有点的线段之终点。每条线段长度为 3 cm
$A \to bA$	与 b 的连接只对该点进行。不允许多连接。b 的方向与 a 相同。b 的长度为0.25 cm。此产生式不能应用超过 10 次
$A \to bB$	a 和 b 的方向必须相同。连接必须简单并只对该点进行此操作
$B \to c$	c 和 a 的方向必须一致。连接必须简单并只对该点进行此操作

作为串识别器的自动机

到现在为止,已经证明文法是模式的生成器。在下边的论述中,考虑识别一个模式是否属于由文法 G 生成的语言 $L(G)$ 的模式。构成句法识别的基本概念可以由称为自动机的计算机器数学模型的发展来说明。给出一个输入模式的串,自动机有能力识别模式是否属于与这个自动机相关联的语言。这里,只关注有限自动机,它是由正则文法生成的语言识别器。

一个有限自动机定义为五元式:

$$A_f = (Q, \Sigma, \delta, q_0, F) \tag{12.3.7}$$

这里 Q 是有限的、非空的状态集合,Σ 是有限的输入字母集合,δ 是从 $Q \times \Sigma$(来自 Q 和 Σ 的元素的有序对构成的集合)到 Q 的所有子集集合的映射,q_0 是起始状态,F(Q 的一个子集)是一个终了或接受状态集合。

例 12.10 一个简单的自动机

考虑由式(12.3.7)给出的自动机,$Q = \{q_0, q_1, q_2\}$,$\Sigma = \{a, b\}$,$F = \{q_0\}$ 及映射 $\delta(q_0, a) = \{q_2\}$,$\delta(q_0, b) = \{q_1\}$,$\delta(q_1, a) = \{q_2\}$,$\delta(q_1, b) = \{q_0\}$,$\delta(q_2, a) = \{q_0\}$ 和 $\delta(q_2, b) = \{q_1\}$。例如,如果自动机处在状态 q_0 并且当前输入为 a,自动机的状态变为 q_2。同样,如果 b 是接下来的输入,自动机的状态变为 q_1,等等,在这种情况下,初始和终结状态是相同的。

图 12.27 显示了刚才谈到的自动机状态图。状态图由表示每个状态的节点和表示可能的状态转换的有向弧组成。终态用双圆圈表示，每条弧用符号标注，它引起由弧连接的状态间的转换。在这种情况下，初始状态和终结状态是相同的。一个终端符号串 w 如果从状态 q_0 开始，在扫描了串 w 的最后一个字符后，字符序列(和 w 被扫描的顺序一样是从左到右读入)使自动机到达终结状态，就称符号串 w 被接受了或被识别了。例如，图 12.27 中的自动机识别串 $w = abbabb$，但拒绝串 $w = aabab$。

正则文法和有限自动机是一一对应的。即，当且仅当它是由一个正则文法生成的，一种语言被有限自动机识别。一个基于刚才谈到概念的句法串识别器的设计是一个直观过程，从一个给定的正则文法得到一个有限自动机。令文法由 $G = (N, \Sigma, P, X_0)$ 表示，这里 $X_0 \equiv S$，并假设 N 由 X_0 加上 n 个附加的非终端符 $X_1, X_2, \cdots, X_n$ 组成。自动机的集合 Q 由引入的 $n+2$ 个状态 $\{q_0, q_1, \cdots, q_n, q_{n+1}\}$ 组成，以便 q_i 和 $X_i (0 \leqslant i \leqslant n)$ 相对应，且 q_{n+1} 为终结状态。

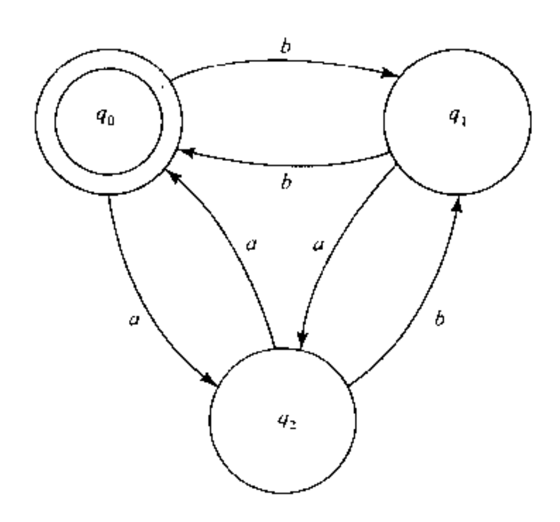


图 12.27　一个有限自动机

输入符号集合与 G 中的终端符集合相一致。δ 中的映射通过使用基于 G 的产生式的两条规则得到；即，对每个 i 和 j，$0 \leqslant i \leqslant n$，$0 \leqslant j \leqslant n$：

1. 如果 $X_i \to aX_j$ 属于 P，则 $\delta(q_i, a)$ 包含 q_j。
2. 如果 $X_i \to a$ 属于 P，则 $\delta(q_i, a)$ 包含 q_{n+1}。

相反，给定一个有限自动机，$A_f = (Q, \Sigma, \delta, q_0, F)$，令 Q 的元素组成 N，起始符 X_0 对应 q_0，G 的产生式用下列方法取得：

1. 如果 q_j 属于 $\delta(q_i, a)$，则 P 中有产生式 $X_i \to aX_j$。
2. 如果 F 中有状态属于 $\delta(q_i, a)$，则 P 中有产生式 $X_i \to a$。

由此得到对应的正则文法，$G = (N, \Sigma, P, X_0)$。两种情况下的终端符号集合是一样的。

例 12.11　识别图 12.26 中模式的有限自动机

与图 12.26 相关的文法的有限自动机，是通过将产生式写为 $X_0 \to aX_1$，$X_1 \to bX_1$，$X_1 \to bX_2$，$X_2 \to c$，然后 $A_f = (Q, \Sigma, \delta, q_0, F)$，$Q = \{q_0, q_1, q_2, q_3\}$，$\Sigma = \{a, b, c\}$，$F = \{q_3\}$ 和映射 $\delta(q_0, a) = \{q_1\}$，$\delta(q_1, b) = \{q_1, q_2\}$，$\delta(q_2, c) = \{q_3\}$ 得到的。为了使其具有完备性，令

$\delta(q_0,b)=\delta(q_0,c)=\delta(q_1,a)=\delta(q_1,c)=\delta(q_2,a)=\delta(q_2,b)=\varnothing$,这里$\varnothing$代表空集,表示本自动机未定义这些转换。

12.3.4 树的句法识别

下面要谈的形式与前面讨论过的串形式相似,现在将讨论扩展到模式的树形描述。假设图像区域或感兴趣的对象通过使用如 11.5 节中所讨论的适当原始基元,以树的形式进行表达。

树文法

树文法以五元式定义:

$$G=(N,\Sigma,P,r,S) \tag{12.3.8}$$

这里,和以前一样,N 和Σ分别为非终端符集合和终端符集合;S 包含于 N 中,是起始符,一般来讲,它也可独立构成一棵树;P 是形如 $T_i \to T_j$ 的产生式集合,这里 T_i 和 T_j 为树;r 是秩函数,表示在文法中节点的直接下降(后代)数目,这些符号在文法中标为一个终端符。所讨论文法的特殊关联性是具有如下形式产生式的开销很大的树文法。

$$X \to \begin{array}{c} k \\ \diagup \ | \ \diagdown \\ X_1 \ X_2 \ldots X_n \end{array}$$

这里,$X_1,X_2,\cdots,X_n$ 为非终端符,k 是终端符。

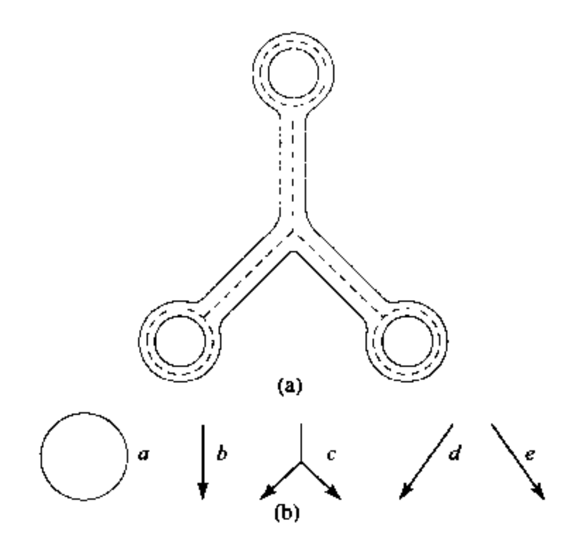
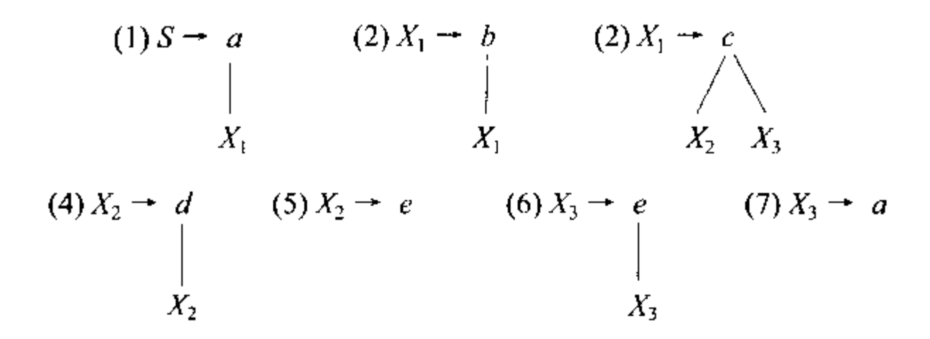


图 12.28　(a)一个对象和(b)借助于树文法表示骨架的基元

例 12.12　一个简单的树文法

图 12.28(a)所示的骨架结构可以使用树文法产生,其中 $N=\{X_1,X_2,X_3,S\}$,$\Sigma=\{a,b,c,d,e\}$,这里终端符代表图 12.28(b)中所示的基元。假设线性基元有首尾的连接性并沿着圆周有任意的连接,在这种考虑下的文法具有如下形式的产生式:



此时的秩函数为 $r(a)=\{0,1\}$，$r(b)=r(d)=\{1\}$，$r(e)=\{0,1\}$，$r(c)=\{2\}$。限制产生式(2)，(4)，(6)应用相同的次数会生成所有3个支路具有相同长度的结构。同样，产生式(4)，(6)应用相同的次数会生成关于垂直轴对称的结构。这种类型的语义信息与先前表12.1相关的讨论相似，并且基于图1.23的知识。

树自动机

尽管传统的有限自动机从左到右逐符号扫描输入字符串，但一个树自动机必须同时从输入树末端(对叶子节点采取从左到右的顺序)的每个节点开始，并沿着并行的路线向着根节点方向处理。特别地，一个末端到根的自动机定义为：

$$A_t = (Q, F, \{f_k \mid k \in \Sigma\}) \tag{12.3.9}$$

这里

Q 是一个有限的状态集合，

F 是 Q 的子集，是一个有限状态集，

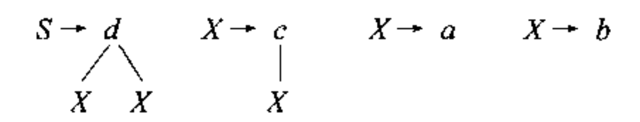
f_k 是 $Q^m \times Q$ 上的关系，m 是 k 的秩。

符号 Q^m 指 Q 的 m 次笛卡儿乘积：$Q^m = Q \times Q \times Q \times \cdots \times Q$。从笛卡儿乘积的定义，我们知道这个表达式意味着带有来自 Q 的元素的有序 m 元组集合。例如，如果 $m=3$，则 $Q^3 = Q \times Q \times Q = \{x, y, z \mid x \in Q, y \in Q, z \in Q\}$。回想从集合 A 到集合 B 的关系 R 是集合 A 和 B 的笛卡儿乘积的子集；即，$R \subseteq A \times B$。因此，在 $Q^m \times Q$ 上的关系仅仅是集合 $Q^m \times Q$ 的一个子集。

对一个开销很大的树文法，$G = (N, \Sigma, P, r, S)$，令 $Q = N$，$F = \{S\}$并对每个属于Σ的字符 a 定义一个关系f_k，以便当且仅当 G 中有产生式：

$$X \to k \quad (\text{子节点 } X_1\ X_2 \ldots X_m)$$

时，$(X_1, X_2, \cdots, X_m, X)$ 在 f_k 内来构造相应的树自动机。例如，考虑树文法 $G = (N, \Sigma\ P, r, S)$，其中 $N = \{S, X\}$，$\Sigma = \{a, b, c, d\}$，产生式：



并且秩为 $r(a)=\{0\}$，$r(b)=\{0\}$，$r(c)=\{1\}$，$r(d)=\{2\}$。相应的树自动机，$A_t=(Q,F,\{f_k \mid k\in\Sigma\})$，通过令 $Q=\{S,X\}$，$F=\{S\}$和$\{f_k \mid k\in\Sigma\}=\{f_a,f_b,f_c,f_d\}$来指定，这里关系定义为：

$f_a=\{(\varnothing,X)\}$，由产生式 $X\rightarrow a$ 产生

$f_b=\{(\varnothing,X)\}$，由产生式 $X\rightarrow b$ 产生

$f_c=\{(X,X)\}$，由产生式 $X\rightarrow$ c
　　　　　　　　　　　　　　　　│
　　　　　　　　　　　　　　　　X

和

$f_d=\{(X,X,S)\}$，由产生式 $S\rightarrow$ d（下接 X X）→产生。

关系 f_a 的解释是，一个标记为 a 的没后代(因此用空符号$\varnothing$)节点赋予状态 X。关系 f_c 的解释是，一个标记为 c 的(带有一个有状态 X 的后代节点)节点赋予状态 X。关系 f_d 的解释是，一个标记为 d 的(带有两个后代节点，每个后代节点都有状态 X)节点赋予状态 S。

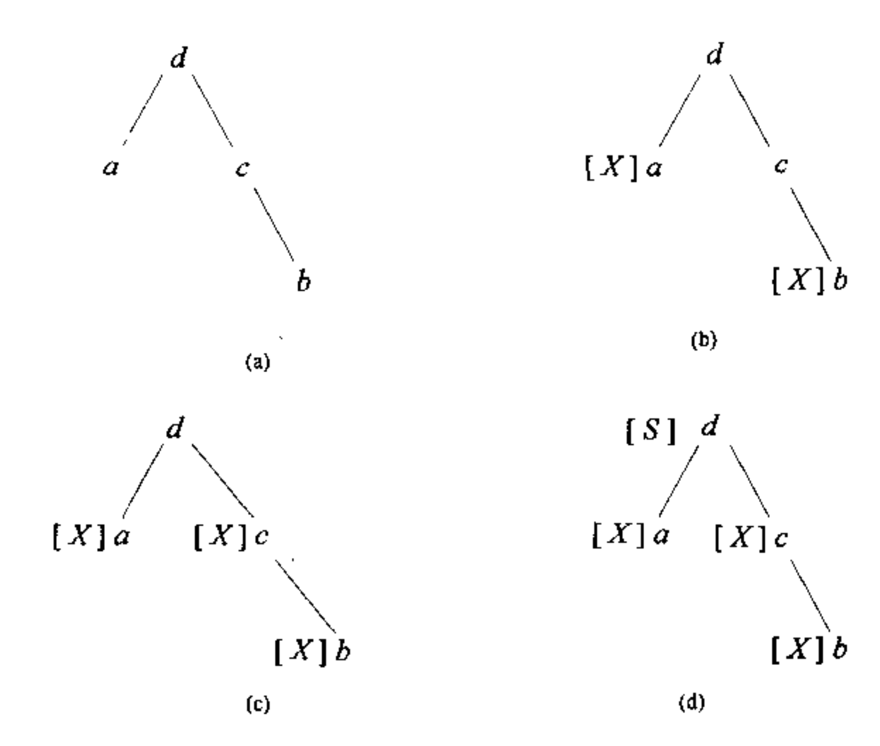


图 12.29　自底向根的树自动机的处理阶段：(a)输入树，(b)赋予叶子节点的状态，(c)赋予中间节点的状态，(d)赋予根节点的状态

为了了解这个树自动机如何识别由前面讨论过的文法生成的树，考虑图 12.29(a)显示的树。自动机 A_t 首先分别通过关系 f_a 和 f_b 将状态赋予末端节点 a 和 b。在这种情况下，根据这两个关系，状态 X 赋给两个叶子节点，如图 12.29(b)所示。现在，自动机从末端节点向上移动一个层次并以 f_c 和节点的后代状态为基础，使一个状态赋予节点 c。再一次基于 f_c 的状态赋

值为 X，如图 12.29(c)所示。再向上移动一层，自动机遇到节点 d，由于状态已经赋予 d 的两个后代，所以要求调用关系 f_d 把状态 S 赋予节点 d。因为这是最后一个节点，并且状态 S 属于 F，自动机接受(识别)此树作为前面提到的树文法所产生语言的有效成员。图 12.29(d)显示了沿着自底向根路径的状态顺序的最终描述图。

例 12.13 使用树文法识别气室中现象的图像

在高能物理实验中气室中发生的现象被记录为一幅图像。这个物理实验是将一束已知性质的粒子导向已知核的目标。典型的现象由从碰撞点发散出来的次级粒子的运行轨迹组成，如图 12.30 显示的例子。引入的轨迹是水平的平行线。注意照片中心附近结果的自然树状结构。

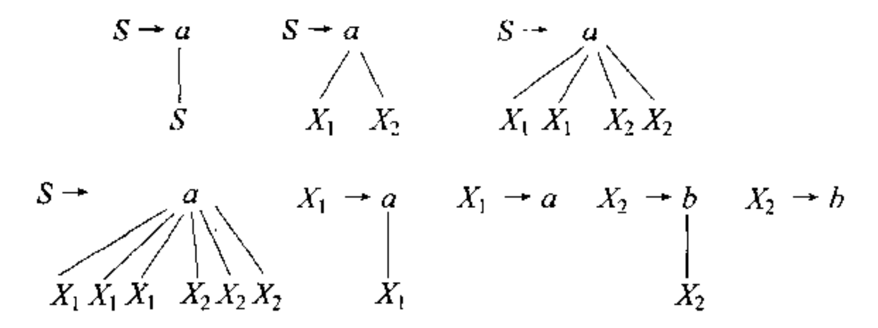
典型的实验产生数以十万计的照片，其中许多没有包含感兴趣的现象。对这些照片进行人工检查和分类费时费力，因此产生了对自动现象识别技术的需求。

一个树文法 $G=(N,\Sigma,P,r,S)$ 可以被指定为生成的这些典型事件的树状描述，代表在氢气室中确定引入带电粒子流后发现的典型现象的树。此时，$N=\{S,X_1,X_2\}$，$\Sigma=\{a,b\}$，元素 a 和 b 有如下解释：

a：⌒ 凸起弧

b：◡ 凹下弧

P 中产生式为：



秩为 $r(a)=\{0,1,2,4,6\}$，$r(b)=\{0,1\}$。分支产生式代表从碰撞点发出的发散轨迹的数目，一般成对出现，通常不超过 6 条。图 12.31(a)显示了图 12.30 中的碰撞现象被分割为凸起和凹下的部分，并且图 12.31(b)显示了对应的树表示。这棵树及其各种变形可以通过上述文法生成。

需要识别刚刚讨论的树类型的树自动机由前边概括的过程来定义。因此有，$A_t=(Q,F,\{f_k \mid k\in\Sigma\})$ 通过令 $Q=\{S,X_1,X_2\}$，$F=\{S\}$ 和 $\{f_k \mid k\in\Sigma\}=\{f_a,f_b\}$ 指定。关系定义为：$f_a=\{(S,S),(X_1,X_2,S),(X_1,X_1,X_2,X_2,S),(X_1,X_1,X_1,X_2,X_2,X_2,S),(X_1,X_1),(\varnothing,X_1)\}$ 且 $f_b=\{(X_2,X_2),(\varnothing,X_2)\}$。我们将这个例子留做练习以显示这个自动机可以接受图 12.31(b)中的树。

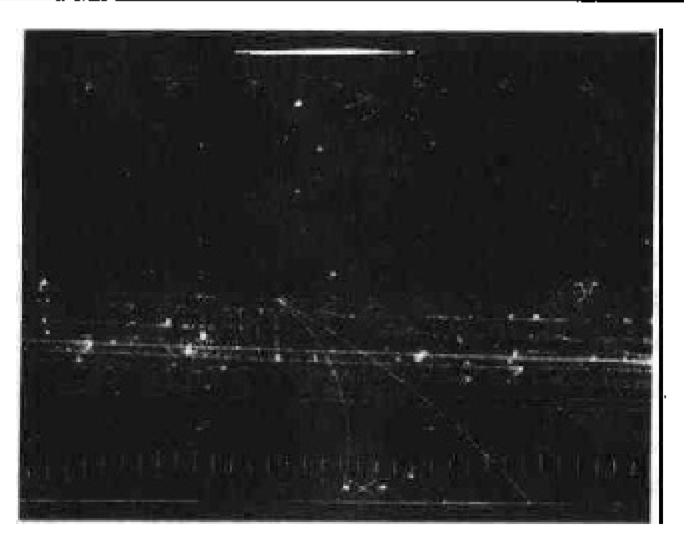


图 12.30 一张气室照片(Fu 和 Bhargava)

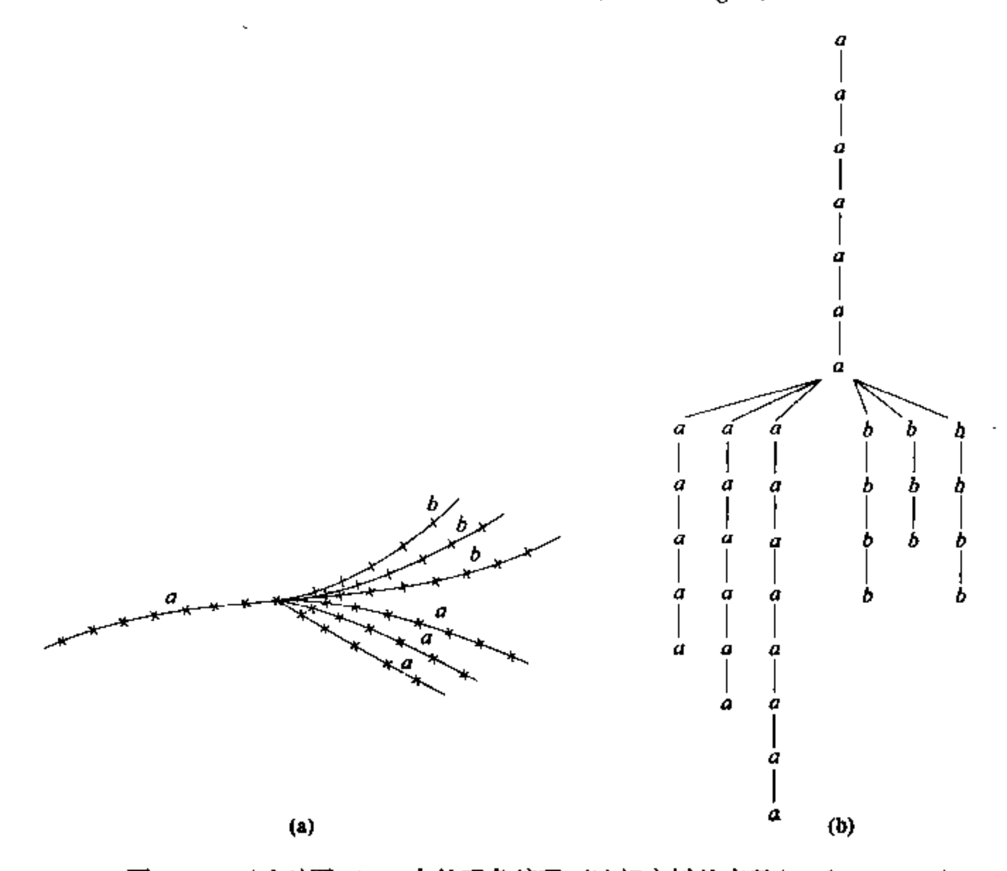


图 12.31 (a)对图 12.30 中的现象编码、(b)相应树的表示(Fu 和 Bhargava)

学习

前段介绍的句法识别方法需要为每个要考虑的类规定一个合适的自动机(识别器)。对于简单的情况,检验可以产生需要的自动机。在更复杂的情况下,自动机从样本模式(如串或树)中学习的算法也是需要的。因为前述的文法和自动机一一对应,直接由样本模式学习文法来提出学习问题,这一过程有时称为文法推论。直接从样本模式串学习有穷自动机是我们关注

的焦点。这一章结尾提供了学习树文法及其自动机和其他句法识别方法的指导。

假如一个类的所有模式由一个未知文法 G 产生，并假设具有下列性质的有限样本集合 R^+

$$R^+ \subseteq \{v \mid v \text{ in } L(G)\} \tag{12.3.10}$$

是可用的。集合 R^+ 称为正样本集。它仅是来自与文法 G 有关的类训练模式的一个集合。如果 G 中每个产生式都被应用于生成至少一个 R^+ 中的元素，则就说这个样本集合在结构上是完善的。我们希望学习(组成)一个有限自动机 A_f，它将接受来自 R^+ 的字串并可能接受类似 R^+ 的字串。

基于有限自动机的定义和 G 与 A_f 的对应性，允许 $R^+ \subseteq \Sigma^*$，这里 Σ^* 是由来自 Σ 的元素组成的所有串的集合。令 Σ^* 中的 z 是一个串，对于 Σ^* 中的某些 w，有 zw 仍属于 R^+。对于一个正整数 k，将关于 R^+ 的 z 的 k 尾定义为集合 $h(z, R^+, k)$，其中：

$$h(z, R^+, k) = \{w \mid zw \text{ in } R^+, |w| \leqslant k\} \tag{12.3.11}$$

换句话说，z 的 k 尾是带有以下性质的串 w 的集合：(1)zw 属于 R^+ 和(2)w 的长度小于或等于 k。

从一个样本集合 R^+ 和一个 k 的特殊值学习自动机 $A_f(R^+, k) = (Q, \Sigma, \delta, q_o, F)$ 的过程，由令：

$$Q = \{q \mid q = h(z, R^+, k)\ ,\ z \text{ in } \Sigma^*\} \tag{12.3.12}$$

和对 Σ 中的每个 a，

$$\delta(q, a) = \{q' \text{in } Q \mid q' = h(za, R^+, k),\ q = h(z, R^+, k)\} \tag{12.3.13}$$

组成。

另外，令：

$$q_0 = h(\lambda, R^+, k) \tag{12.3.14}$$

且

$$F = \{q \mid q \text{ in } Q, \lambda \text{ in } q\} \tag{12.3.15}$$

这里，λ 是空串(没有字符的串)。注意到自动机 $A_f(R^+, k)$ 具有从 R^+ 构造的所有 k 尾集合的状态子集。

例 12.14 推导一个来自样本模式的有限自动机

假如 $R^+ = \{a, ab, abb\}$ 且 $k = 1$。则从前边的讨论有：

$$
\begin{aligned}
z = \lambda, \qquad & h(\lambda, R^+, 1) = \{w \mid \lambda w \text{ in } R^+, |w| \leqslant 1\} \\
& \qquad = \{a\} \\
& \qquad = q_0 \\
z = a, \qquad & h(a, R^+, 1) = \{w \mid aw \text{ in } R^+, |w| \leqslant 1\} \\
& \qquad = \{\lambda, b\} \\
& \qquad = q_1 \\
z = ab, \qquad & h(ab, R^+, 1) = \{\lambda, b\} \\
& \qquad = q_1
\end{aligned}
$$

$$z = abb, \qquad h(abb,R^+,1) = \{\lambda\}$$
$$= q_2$$

在这种情况下,$\sum^*$ 中的其他串 z 生成不属于 R^+ 的串 zw,使自动机到达第四个状态,由 $q_\varnothing$ 定义,这与 h 为空集的状态相对应。由此,状态为:$q_0 = \{a\}$,$q_1 = \{\lambda, a\}$,$q_2 = \{\lambda\}$ 和 $q_\varnothing$,它们给出集合 $Q = \{q_0, q_1, q_2, q_\varnothing\}$。尽管状态由字符集合($k$ 尾数)得到,状态符号 $q_0, q_1, \cdots$ 仅仅被用于形成集合 Q。下一步是得到转移函数。由于 $q_0 = h(\lambda, R^+, 1)$,得出:

$$\delta(q_0, a) = h(\lambda a, R^+, 1) = h(a, R^+, 1) = q_1$$

和

$$\delta(q_0, b) = h(\lambda b, R^+, 1) = h(b, R^+, 1) = q_\varnothing$$

类似地,

$$q_1 = h(a, R^+, 1) = h(ab, R^+, 1)$$

得出:

$$\delta(q_1, a) = h(aa, R^+, 1) = h(aba, R^+, 1) = q_\varnothing$$

还有,$\delta(q_1, b) \supseteq h(ab, R^+, 1) = q_1$ 且 $\delta(q_1, b) \supseteq h(abb, R^+, 1) = q_2$,即 $\delta(q_1, b) = \{q_1, q_2\}$。按照刚才的描述,得出 $\delta(q_2, a) = \delta(q_2, b) = \delta(q_\varnothing, a) = \delta(q_\varnothing, b) = q_\varnothing$。终结状态集合包含那些在 k 尾中有空字符串 λ 的状态。此时,$q_1 = \{\lambda, b\}$ 且 $q_2 = \{\lambda\}$,所以 $F = \{q_1, q_2\}$。

基于这些结果,推导的自动机给出如下:

$$A_f(R^+, 1) = (Q, \sum, \delta, q_0, F)$$

这里,$Q = \{q_0, q_1, q_2, q_\varnothing\}$,$\sum = \{a, b\}$,$F = \{q_1, q_2\}$,转移函数如上面给出的。图 12.32 显示了状态图。自动机接受形如 $a, ab, abb, \cdots, ab^n$ 的串,这些串和给定的样本集一致。

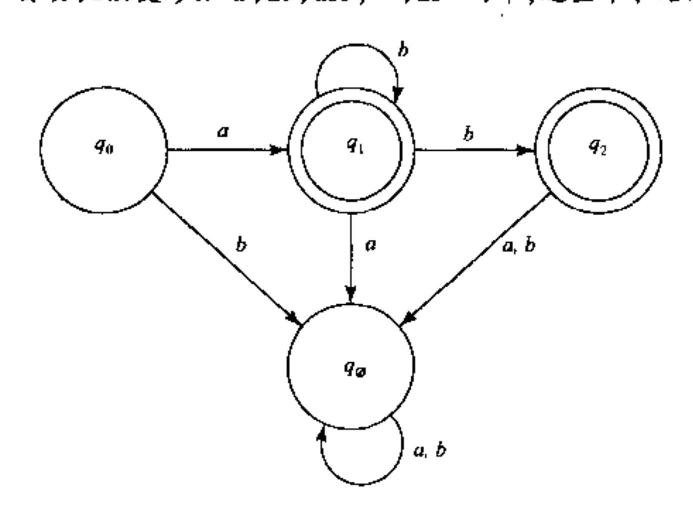


图 12.32 由样本集合 $R^+ = \{a, ab, abb\}$ 推出的有限自动机状态图

前述例子显示出 k 值控制着结果自动机的性质。下面的性质说明 $A_f(R^+,k)$ 对这一参数的依赖性。

性质 1：对于所有的 $k \geq 0$，$R^+ \subseteq L[A_f(R^+,k)]$，这里 $L[A_f(R^+,k)]$ 是 $A_f(R^+,k)$ 接受的语言。

性质 2：如果 k 大于或等于 R^+ 中最长的串长度，则 $L[A_f(R^+,k)] = R^+$；如果 $k=0$，则 $L[A_f(R^+,k)] = \Sigma^*$。

性质 3：$L[A_f(R^+,k+1)] \subseteq L[A_f(R^+,k)]$

性质 1 保证 $A_f(R^+,k)$ 作为最小值，将接受样本集合 R^+ 中的串。如果 k 大于或等于 R^+ 中最长串的长度，则由性质 2，自动机将仅接受 R^+ 中的串。如果 $k=0$，$A_f(R^+,0)$ 将由一个状态 $q_0=\{\lambda\}$ 构成，它既是初始状态又是终结状态。对 Σ 中的 a，转移函数将具有形式 $\delta(q_0,a)=q_0$。所以，$L[A_f(R^+,0)]=\Sigma^*$，并且自动机将接受空串 λ 和所有由来自 Σ 的字符组成的串。最后，性质 3 指出 $A_f(R^+,k)$ 接受的语言规模在 k 增大时减小。

这三条性质允许仅仅通过改变参数 k 来控制 $A_f(R^+,k)$ 的性质。如果 $L[A_f(R^+,k)]$ 是来自被选中样本 R^+ 的语言 L_0 的一种猜测，并且 k 值非常小，那么这个语言 L_0 的猜测会构成一个也许包含 Σ^* 中大部分或全部串的自由推论。然而，如果 k 等于 R^+ 中最长串的长度，$A_f(R^+,k)$ 将仅接受包含在 R^+ 中串的推论会比较保守。图 12.33 用图形显示了这些概念。

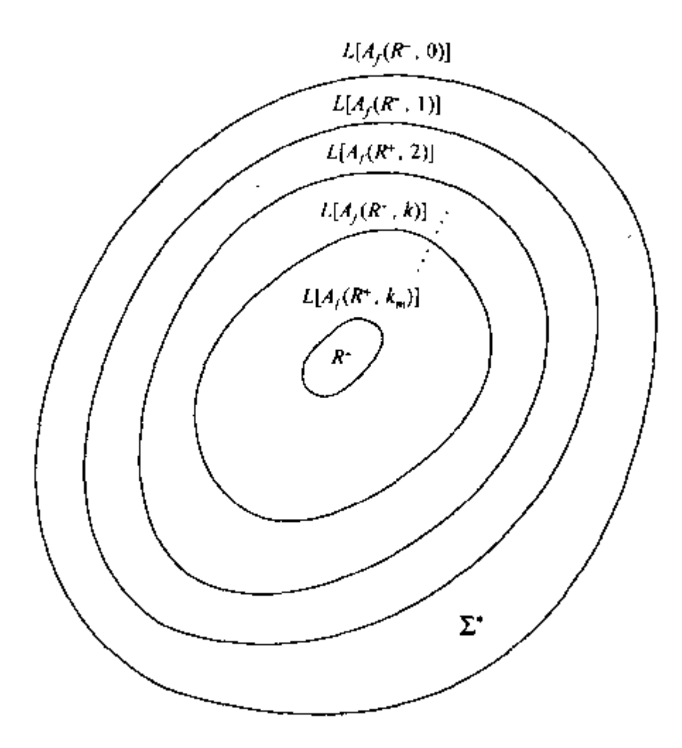


图 12.33　$L[A_f(R^+,k)]$ 和 k 的关系，值 k_m 大于或等于 R^+ 中最长串的长度

例 12.15　从给定模式集合中推导自动机的另一个例子

考虑集合 $R^+=\{caaab, bbaab, caab, bbab, cab, bbb, cb\}$。对 $k=1$，使用前述方法得出下列相同的过程：

1. $z=\lambda$, $h(\lambda,R^+,1)=\{\varnothing\}=q_{\varnothing}$;
2. $z=c$, $h(z,R^+,1)=\{b\}=q_1$;
3. $z=ca$, $h(z,R^+,1)=\{b\}=q_1$;
4. $z=cb$, $h(z,R^+,1)=\{\lambda\}=q_0$;
5. $z=caa$, $h(z,R^+,1)=\{b\}=q_1$;
6. $z=cab$, $h(z,R^+,1)=\{\lambda\}=q_0$;
7. $z=caaa$, $h(z,R^+,1)=\{b\}=q_1$;
8. $z=caab$, $h(z,R^+,1)=\{\lambda\}=q_0$;
9. $z=caaab$, $h(z,R^+,1)=\{\lambda\}=q_0$;
10. $z=b$, $h(z,R^+,1)=\{\varnothing\}=q_{\varnothing}$;
11. $z=bb$, $h(z,R^+,1)=\{b\}=q_1$;
12. $z=bba$, $h(z,R^+,1)=\{b\}=q_1$;
13. $z=bbb$, $h(z,R^+,1)=\{\lambda\}=q_0$;
14. $z=bbaa$, $h(z,R^+,1)=\{b\}=q_1$;
15. $z=bbab$, $h(z,R^+,1)=\{\lambda\}=q_0$;
16. $z=bbaab$, $h(z,R^+,1)=\{\lambda\}=q_0$。

自动机为：

$$A_f(R^+,1)=(Q,\Sigma,\delta,q_0,F)$$

$Q=\{q_0,q_1,q_{\varnothing}\}$，$\Sigma=\{a,b,c\}$，$F=\{q_0\}$，状态转移示于图 12.34 中的状态图。为了被自动机接受，串必须以 a，b 或 c 开始，并以 b 结尾。同样，带 a，b 或 c 的循环串由 $A_f(R^+,1)$ 接受。

前述方法的主要优点是实现简单。通过适当的努力，综合过程可以在数字计算机上模拟。主要的缺点是，尽管问题通过上述三条性质简化到一定程度，但还是需要决定恰当的 k 值。

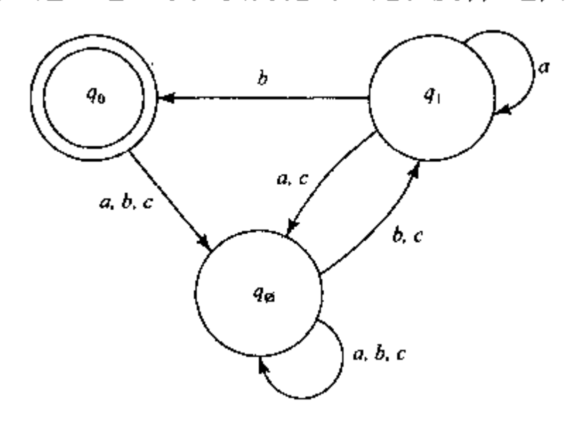


图 12.34 从样本集合 $R^+=\{caaab, bbaab, caab, bbab, cab, bbb, cb\}$ 推导出来的自动机 $A_f(R^+,1)$ 的状态图

小结

从第 9 章开始，对数字图像处理的论述发生了转变，从论述输出图像的过程变为论述输出图像特性的过程，在 1.1 节中进行了详细解释。本章中的材料本质上是介绍性的，覆盖的题目对于理解对象识别技术的状态很基础。如在本章开头提到的，单个对象的识别是本书结论合乎逻辑的安排。超越这一点，就需要在 1.4 节中给出的范围以外的概念。特别地，接下来合乎逻辑的步骤将是图像分析方法的研究，这一发展需要机器智能的一些概念。

如在 1.1 节和 1.4 中提到的，机器智能及依赖于它的某些领域，比如场景分析和计算机视觉，仍处在它们实践发展的初期阶段。今天，图像分析问题的解决仍是通过具有探索性的方法进行的。当这些方法的确发生了变化时，它们中的大部分恰恰是应用了本书中所涵盖的重要基础技术。

在总结研究了前述 12 章中的材料后，无论从理论上还是从实践上，读者现在应该可以领略到数字图像处理领域的主要内容了。对于所有的论述请细加琢磨，以便为以后进一步研究本领域或研究以此为基础的相关领域打下坚实的基础。利用许多成像问题特定任务的性质，对基本原理的清晰理解可以显著增加成功解决这些问题的可能性。

参考资料

12.1 节到 12.2.2 节的背景材料来自 Duda，Hart 和 Stork[2001]，还有 Tou 和 Gonzalez[1974]的书籍。由 Jain 等人[2000]写的综述文章也很重要。Principe 等人[1999]所写的书对神经网络进行了很好的描述。电气和电子工程师协会论文集(IEEE Trans.)的特刊“图像处理”[1998]是可以同十年前的一个特刊(IEEE Computer[1988])相媲美的刊物。12.2.3 节中提到的材料是介绍性的。实际上，那段论述使用的神经网络模型是多年来提出的大量模型中的一个。然而，我们论述的这个模型是有代表性的，而且是在图像处理过程中使用得相当广泛的模型。识别失真图形的例子改编了 Gupta 等人[1990，1994]的例子。Gori 和 Scarselli[1998]写的论文论述了多层次神经网络的分类能力。由 Ueda[2000]报告的一个基于使用线性连接的神经网络的方法使分类错误降至最低，这是文中另外一个很好的阅读材料。

12.3.1 节中其他附加的阅读材料见 Bribiesca 和 Guzman[1980]的著述。在串匹配方面，见 Sze 和 Yang[1981]，Oommen 和 Loke[1997]，还有 Gdalyahu 和 Weinshall[1999]的论著。12.3.3 节和 12.3.4 节参考了 Gonzalez 和 Thomason[1978]，Fu[1982]，还有 Bunke 和 Sanfeliu[1990]的著述。也可以参见 Tanaka[1995]，Vailaya 等人[1998]，Aizaka 和 Nakamura[1999]，还有 Jonk 等人[1999]的著述。

习题

12.1　(a)对图 12.1 中所示模式计算最小距离分类器的判别函数。通过检测(要仔细检测)也许能得到需要的均值向量。

(b)由(a)得到的判别函数绘制决策平面。

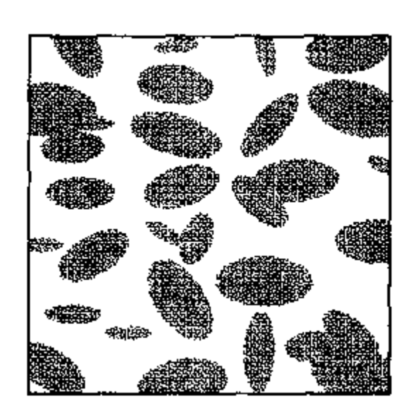
★12.2 说明在模式分类方面,式(12.2.4)和式(12.2.5)具有同样的功能。

12.3 说明由式(12.2.6)给出的平面是 n 维点 $\mathbf{m}_i$ 和 $\mathbf{m}_j$ 间连线的垂直等分线。

★12.4 联系图 12.7 的讨论,说明最小距离分类器如何使用 W 电阻排(W 是类别数)、对每排电阻进行的相加连接(将电流加起来)和一个能选择 W 最大值的最大值选择器来实现。

12.5 说明式(12.2.8)的相关系数具有[-1,1]范围内的值。
提示:用矢量形式表示 $\gamma(x,y)$。

★12.6 一个实验产生带有斑点的二值图像,这些小斑点外形近似椭圆(见下图)。斑点有 3 种尺寸,椭圆的主轴平均值为(1.3,0.7),(1.0,0.5),(0.75,0.25)。这些轴的尺寸相对于它们的平均值有 ±10% 的变化。开发一种图像处理系统,它能够拒绝不完整的或相互重叠的椭圆并依据给出的 3 种尺寸对剩下的单个椭圆进行分类。以方框图的形式显示结果,对于每个部分的操作给出详尽的细节。使用最小距离分类器解决分类问题,清楚地指出如何得到训练样本和如何使用这些样本训练分类器。



12.7 下列模式类具有高斯概率密度函数 $\omega_1:\{(0,0)^T,(2,0)^T,(2,2)^T,(0,2)^T\}$ 和 $\omega_2:\{(4,4)^T,(6,4)^T,(6,6)^T,(4,6)^T\}$。

(a)假设 $P(\omega_1)=P(\omega_2)=\frac{1}{2}$,求两个类之间的贝叶斯决策边界方程式。

(b)画出边界线。

★12.8 重复习题 12.7,但使用下列模式类:$\omega_1:\{(-1,0)^T,(0,-1)^T,(1,0)^T,(0,1)^T\}$ 和 $\omega_2:\{(-2,0)^T,(0,-2)^T,(2,0)^T,(0,2)^T\}$。显然这些类不是线性可分的。

12.9 重复习题 12.6,但使用贝叶斯分类器(假设有高斯密度)。清楚地指出如何得到训练样本和如何使用这些样本训练分类器。

★12.10 贝叶斯判别函数 $d_j(\mathbf{x})=p(\mathbf{x}/\omega_j)P(\omega_j)$,$j=1,2,\cdots,W$,用 0-1 失效函数推导。证明这些决策函数最小化了出错概率[提示:出错概率 $p(e)$ 是 $1-p(c)$,这里 $p(c)$ 是正确分类的概率]。对于一个属于类 ω_i 的模式向量 $\mathbf{x}$,$p(c/\mathbf{x})=p(\omega_i/\mathbf{x})$。找出 $p(c)$ 并说明当 $p(\mathbf{x}/\omega_i)P(\omega_i)$ 是最大值时,$p(c)$ 取最大值[$p(e)$ 为最小值]。

12.11 (a)对下列模式类使用感知器算法:$\omega_1:\{(0,0,0)^T,(1,0,0)^T,(1,0,1)^T,(1,1,0)^T\}$ 和 $\omega_2:\{(0,0,1)^T,(0,1,1)^T,(0,1,0)^T,(1,1,1)^T\}$。令 $c=1$ 和 $\mathbf{w}(1)=$

$(-1,-2,-2,0)^T$。

(**b**)画出(a)中得到的决策平面。说明模式类并指出决策面正的一侧。

★12.12 式(12.2.34)到式(12.2.36)给出的感知器算法可以通过用 -1 乘以模式类 ω_2 以表达得更为简练。此时,算法的正确步骤变为:如果 $\mathbf{w}^T(k)\mathbf{y}(k)>0$,则$\mathbf{w}(k+1)=\mathbf{w}(k)$,否则$\mathbf{w}(k+1)=\mathbf{w}(k)+c\mathbf{y}(k)$。这是几个由普通梯度下降公式(见下式)导出的感知器算法公式之一。

$$\mathbf{w}(k+1)=\mathbf{w}(k)-c\left[\frac{\partial J(\mathbf{w},\mathbf{y})}{\partial \mathbf{w}}\right]_{\mathbf{w}=\mathbf{w}(k)}$$

这里 $c>0$, $J(\mathbf{w},\mathbf{y})$是准则函数,偏导数为 $\mathbf{w}=\mathbf{w}(k)$处的值。说明通过使用准则函数 $J(\mathbf{w},\mathbf{y})=\frac{1}{2}(|\mathbf{w}^T\mathbf{y}|-\mathbf{w}^T\mathbf{y})$,感知器算法公式可以在普通梯度降低过程中得到。这里 $|\arg|$ 是变元的绝对值(注意:关于 $\mathbf{w}$ 的 $\mathbf{w}^T\mathbf{y}$ 偏导数等于 $\mathbf{y}$)。

12.13 证明如果训练模式集合是线性可分离的,在式(12.2.34)到式(12.2.36)中给出的感知器训练算法收敛于有限步[提示:将模式类 ω_2 乘以 -1,并考虑非负门限 T,使感知器训练算法($c=1$)表示为:如果 $\mathbf{w}^T(k)\mathbf{y}(k)>T$,则 $\mathbf{w}(k+1)=\mathbf{w}(k)$,否则 $\mathbf{w}(k+1)=\mathbf{w}(k)+\mathbf{y}(k)$。也许要利用柯西-施瓦茨不等式:$\|\mathbf{a}\|^2\|\mathbf{b}\|^2\geqslant(\mathbf{a}^T\mathbf{b})^2$]。

★12.14 详细说明神经网络的结构和权值可以完全实现与 n 维空间中对两个模式类分类的最小距离分类器相同的功能。

12.15 详细说明神经网络的结构和权值可以完全实现与 n 维空间中对两个模式类分类的贝叶斯分类器相同的功能。这些类都是高斯的并有不同的均值,但其协方差矩阵是相等的。

★12.16 (**a**)在什么条件下,习题 12.14 和习题 12.15 中的神经网络一样?

(**b**)如果用足够数量的样本进行训练,12.2.3 节中开发的多层前馈神经网络产生的德尔塔规则会生成(a)中提到的特殊神经网络吗?

12.17 在两个维度上的两个模式类以下列形式分布:模式类 ω_1 随机地沿着一个半径为 r_1 的圆分布,同样,模式类 ω_2 随机地沿着半径为 r_2 的圆分布,这里 $r_2=2r_1$。用最少的层次和节点指定一个神经网络的结构,对这两个类的模式进行适当分类。

★12.18 重复习题 12.6,但使用神经网络。清楚地指出如何得到训练样本和如何使用这些样本训练分类器。以你的观点选择能解决问题的可能的最简单神经网络。

12.19 说明式(12.2.71)给出的表达式 $h'_j(I_j)=O_j(1-O_j)$[这里 $h'_j(I_j)=\partial h_j(I_j)/\partial I_j$],根据式(12.2.50)($\theta_o=1$)得出。

★12.20 说明式(12.3.2)中的距离度量 $D(A,B)$满足式(12.3.3)的性质。

12.21 说明当且仅当 a 和 b 是同样的串时,式(12.3.4)中的 $\beta=\max(|a|,|b|)-\alpha$ 为0。

12.22 ★(**a**)详细说明一个能够识别形如 ab^na 的模式串的有限自动机。

(**b**)由(a)中的解得到对应的正则文法(不通过检测解决)。

12.23 在两个空间方向上,对 0 和 1 交错组成(类似棋盘模式)的图像给出一个开销大的树文法。假设左上边的元素是 1,并且所有图像结束于左下边的元素 1。

★12.24 使用式(12.3.12)到式(12.3.15)规定的学习过程,去学习一个能识别形如 ab^na 的

串的有限自动机,其中 $n>0$。从样本集合$\{aba, abba, abbba\}$开始。如果这个集合对于使用算法找到字符 b 重复出现的规律是不充足的,则增加样本串,直到算法可以找到其中的规律。

12.25 说明与图 12.30 有关的树自动机可以接受图 12.31(b)中给出的树。

12.26 某一工厂为运动会生产了大量的小美国国旗。质量控制小组发现,在生产高峰期,某些印刷机器有少印(随机地)一到三个星或一到两条整条条纹的倾向。除了这些错误以外,旗子的其他方面都很好。尽管对于全部产品来说,含有错误的旗子只占很少比率,工厂经理仍决定解决这一问题。在大量的调查之后,他总结出应用图像处理技术自动检测是解决问题最经济的方法。具体说明如下:旗子尺寸约为 12.5 cm × 7.5 cm。他们将生产线(独立的,但定位上有 15%的偏差)纵向移动的速度减少到大约 50 cm/s,旗子之间有大约 5 cm 的间隙。这里"大约"指 ± 5% 。工厂经理让你为每条生产线设计一个图像处理系统。在决定所采用方法的可用性时,简单性和花费是重要参量。基于图 1.23 设计一个完善的系统。将你的解决方案简要(但要清楚)写成文档(包括各种假设和说明)向工厂经理汇报。

参考文献

Abidi, M. A., and Gonzalez, R. C. (eds.) [1992]. *Data Fusion in Robotics and Machine Intelligence*, Academic Press, New York.

Abidi, M. A., Eason, R. O., and Gonzalez, R. C. [1991]. "Autonomous Robotics Inspection and Manipulation Using Multisensor Feedback." *IEEE Computer*, vol. 24, no. 4, pp. 17 – 31.

Abramson, N. [1963]. *Information Theory and Coding*, McGraw-Hill, New York.

Adiv, G. [1985]. "Determining Three-Dimensional Motion and Structure from Optical Flow Generated by Several Moving Objects." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-7, no. 4, pp. 384 – 401.

Aggarwal, J. K., and Badler, N. I. (eds.) [1980]. "Motion and Time-Varying Imagery." *IEEE Trans. Pattern Anal. Mach. Intell.*, Special Issue, vol. PAMI-2, no. 6, pp. 493 – 588.

Aguado, A. S., Nixon, M. S., and Montiel, M. M. [1998], "Parameterizing Arbitrary Shapes via Fourier Descriptors for Evidence-Gathering Extraction," *Computer Vision and Image Understanding*, vol. 69, no. 2, pp. 202 – 221.

Ahmed, N., Natarajan, T., and Rao, K. R. [1974]. "Discrete Cosine Transforms." *IEEE Trans. Comp.*, vol. C-23, pp. 90 – 93.

Ahmed, N., and Rao, K. R. [1975]. *Orthogonal Transforms for Digital Signal Processing*, Springer-Verlag, New York.

Aizaka, K., and Nakamura, A. [1999]. "Parsing of Two-Dimensional Images Represented by Quadtree Adjoining Grammars," *Pattern Recog.*, vol. 32, no. 2, pp. 277 – 294.

Alliney, S. [1993]. "Digital Analysis of Rotated Images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, no. 5, pp. 499 – 504.

Ando, S. [2000]. "Consistent Gradient Operators," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 3, pp. 252 – 265.

Andrews, H. C. [1970]. *Computer Techniques in Image Processing*, Academic Press, New York.

Andrews, H. C., and Hunt, B. R. [1977]. *Digital Image Restoration*, Prentice Hall, Englewood Cliffs, N.J.

Anelli, G., Broggi, A., and Destri, G. [1998]. "Decomposition of Arbitrarily-Shaped Morphological Structuring Elements Using Genetic Algorithms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 2, pp. 217 – 224.

Ang, P. H., Ruetz, P. A., and Auld, D. [1991]. "Video Compression Makes Big Gains." *IEEE Spectrum*, vol. 28, no. 10, pp. 16 – 19.

Antonini, M., Barlaud, M., Mathieu, P., and Daubechies, I. [1992]. "Image Coding Using Wavelet Transform." *IEEE Trans. Image Processing*, vol. 1, no. 2, pp. 205 – 220.

Atchison, D. A., and Smith, G. [2000]. *Optics of the Human Eye*, Butterworth-Heinemann, Bos-

ton, Mass.

Baccar, M., Gee, L. A., Abidi, M. A., and Gonzalez, R. C. [1996]. "Segmentation of Range Images Via Data Fusion and Morphological Watersheds," *Pattern Recog.*, vol. 29, no. 10, pp. 1671 – 1685.

Bajcsy, R., and Lieberman, L. [1976]. "Texture Gradient as a Depth Cue." *Comput. Graph. Image Proc.*, vol. 5, no. 1, pp. 52 – 67.

Bakir, T., and Reeves, J. S. [2000]. "A Filter Design Method for Minimizing Ringing in a Region of Interest in MR Spectroscopic Images," *IEEE Trans. Medical Imaging*, vol. 19, no. 6, pp. 585 – 600.

Ballard, D. H. [1981]. "Generalizing the Hough Transform to Detect Arbitrary Shapes." *Pattern Recognition*, vol. 13, no. 2, pp. 111 – 122.

Ballard, D. H., and Brown, C. M. [1982]. *Computer Vision*, Prentice Hall, Englewood Cliffs, N.J.

Banham, M. R., Galatsanos, H. L., Gonzalez, H. L., and Katsaggelos, A. K. [1994]. "Multichannel Restoration of Single Channel Images Using a Wavelet-Based Subband Decomposition." *IEEE Trans. Image Processing*, vol. 3, no. 6, pp. 821 – 833.

Banham, M. R., and Katsaggelos, A. K. [1996]. "Spatially Adaptive Wavelet-Based Multiscale Image Restoration." *IEEE Trans. Image Processing*, vol. 5, no. 5, pp. 619 – 634.

Basart, J. P., and Gonzalez, R. C. [1992]. "Binary Morphology," in *Advances in Image Analysis*, Y. Mahdavieh and R. C. Gonzalez (eds.), SPIE Press, Bellingham, Wash., pp. 277 – 305.

Basart, J. P., Chacklackal, M. S., and Gonzalez, R. C. [1992]. "Introduction to Gray-Scale Morphology," in *Advances in Image Analysis*, Y. Mahdavieh and R. C. Gonzalez (eds.), SPIE Press, Bellingham, Wash., pp. 306 – 354.

Bates, R. H. T., and McDonnell, M. J. [1986]. *Image Restoration and Reconstruction*. Oxford University Press, New York.

Battle, G. [1987]. "A Block Spin Construction of Ondelettes. Part I: Lemarié Functions." *Commun. Math. Phys.*, vol. 110, pp. 601 – 615.

Battle, G. [1988]. "A Block Spin Construction of Ondelettes. Part II: the QFT Connection." *Commun. Math. Phys.*, vol. 114, pp. 93 – 102.

Baumert, L. D., Golomb, S. W., and Hall, M., Jr. [1962]. "Discovery of a Hadamard Matrix of Order 92." *Bull. Am. Math. Soc.*, vol. 68, pp. 237 – 238.

Baxes, G. A. [1994]. *Digital Image Processing: Principles and Applications*, John Wiley & Sons, New York.

Baylon, D. M., and Lim, J. S. [1990]. "Transform/Subband Analysis and Synthesis of Signals." *Tech. Report, MIT Research Laboratory of Electronics. Cambridge. Mass.*

Bell, E. T. [1965]. *Men of Mathematics*, Simon & Schuster, New York.

Bengtsson, A., and Eklundh, J. O. [1991]. "Shape Representation by Multiscale Contour Approximation." *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 1, pp. 85 – 93.

Benson, K. B. [1985]. *Television Engineering Handbook*, McGraw-Hill, New York.

Berger, T. [1971]. *Rate Distortion Theory*, Prentice Hall, Englewood Cliffs, N.J.

Beucher, S. [1990]. Doctoral Thesis, Centre de Morphologie Mathématique, École des Mines de Paris, France. (The core of this material is contained in the following paper.)

Beucher, S., and Meyer, F. [1992]. "The Morphological Approach of Segmentation: The Watershed Transformation," in *Mathematical Morphology in Image Processing*, E. Dougherty (ed.), Marcel Dekker, New York.

Bhaskaran, V., and Konstantinos, K. [1997]. *Image and Video Compression Standards: Algorithms and Architectures*. Kluwer, Boston, Mass.

Bhatt, B., Birks, D., Hermreck, D. [1997]. "Digital Television: Making It Work." *IEEE Spectrum*, vol. 34, no. 10, pp. 19–28.

Biberman, L. M. [1973]. "Image Quality." In *Perception of Displayed Information*. Biberman, L. M. (ed.), Plenum Press, New York.

Bichsel, M. [1998]. "Analyzing a Scene's Picture Set under Varying Lightning," *Computer Vision and Image Understanding*, vol. 71, no. 3, pp. 271–280.

Bieniek, A., and Moga, A. [2000]. "An Efficient Watershed Algorithm Based on Connected Components," *Pattern Recog.*, vol. 33, no. 6, pp. 907–916.

Bisignani, W. T., Richards, G. P., and Whelan, J. W. [1966]. "The Improved Grey Scale and Coarse-Fine PCM Systems: Two New Digital TV Bandwidth Reduction Techniques." *Proc.* IEEE, vol. 54, no. 3, pp. 376–390.

Blahut, R. E. [1987]. *Principles and Practice of Information Theory*, Addison-Wesley, Reading, Mass.

Bleau, A. and Leon, L. J. [2000]. "Watershed-Based Segmentation and Region Merging," *Computer Vision and Image Understanding*, vol. 77, no. 3, pp. 317–370.

Blouke, M. M., Sampat, N., and Canosa, J. [2001]. *Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications-II*, ISBN 0–8194–3583–X, SPIE Press, Bellingham, Wash.

Blum, H. [1967]. "A Transformation for Extracting New Descriptors of Shape." In *Models for the Perception of Speech and Visual Form*, Wathen-Dunn, W. (ed.), MIT Press, Cambridge, Mass.

Blume, H., and Fand, A. [1989]. "Reversible and Irreversible Image Data Compression Using the S-Transform and Lempel-Ziv Coding." *Proc. SPIE Medical Imaging III: Image Capture and Display*, vol. 1091, pp. 2–18.

Boie, R. A., and Cox, I. J. [1992]. "An Analysis of Camera Noise," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 6, pp. 671–674.

Born, M., and Wolf, E. [1999]. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 7th ed., Cambridge University Press, Cambride, UK.

Boulgouris, N. V., Tzovaras, D., and Strintzis, M. G. [2001]. "Lossless Image Compression Based on Optimal Prediction, Adaptive Lifting, and Conditional Arithmetic Coding," *IEEE Trans.*

Image Processing, vol. 10, no. 1, pp. 1 – 14.

Bouman, C., and Liu, B. [1991]. "Multiple Resolution Segmentation of Textured Images." *IEEE Trans. Pattern. Anal. Machine Intell.*, vol. 13, no. 2, pp. 99 – 113.

Boyd, J. E. and Meloche, J. [1998]. "Binary Restoration of Thin Objects in Multidimen-sional Imagery," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 6, pp. 647 – 651.

Bracewell, R. N. [1995]. *Two-Dimensional Imaging*, Prentice Hall, Upper Saddle River, N.J.

Bracewell, R. N. [2000]. *The Fourier Transform and its Applications*, 3rd ed. McGraw-Hill, New York.

Bribiesca, E. [1981]. "Arithmetic Operations Among Shapes Using Shape Numbers." *Pattern Recog.*, vol. 13, no. 2, pp. 123 – 138.

Bribiesca, E. [1999]. "A New Chain Code," *Pattern Recog.*, vol. 32, no. 2, pp. 235 – 251.

Bribiesca, E. [2000]. "A Chain Code for Representing 3-D Curves," *Pattern Recog.*, vol. 33, no. 5, pp. 755 – 765.

Bribiesca, E., and Guzman, A. [1980]. "How to Describe Pure Form and How to Measure Differences in Shape Using Shape Numbers." *Pattern Recog.*, vol. 12, no. 2, pp. 101 – 112.

Brigham, E. O. [1988]. *The Fast Fourier Transform and its Applications*, Prentice Hall, Upper Saddle River, N.J.

Brinkman, B. H., Manduca, A., and Robb, R. A. [1998]. "Optimized Homomorphic Unsharp Masking for MR Grayscale Inhomogeneity Correction," *IEEE Trans. Medical Imaging*, vol. 17, no. 2, pp. 161 – 171.

Brummer, M. E. [1991]. "Hough Transform Detection of the Longitudinal Fissure in Tomographic Head Images." *IEEE Trans. Biomed. Images*, vol. 10, no. 1, pp. 74 – 83.

Brzakovic, D., Patton, R., and Wang, R. [1991]. "Rule-Based Multi-template Edge Detection." *Comput. Vision, Graphics, Image Proc: Graphical Models and Image Proc.*, vol. 53, no. 3, pp. 258 – 268.

Bunke, H. and Sanfeliu, A. (eds.) [1990]. *Syntactic and Structural Pattern Recognition: Theory and Applications*, World Scientific, Teaneck, N.J.

Burrus, C. S., Gopinath, R. A., and Guo, H. [1998]. *Introduction to Wavelets and Wavelet Transforms*, Prentice Hall, Upper Saddle River, N J., pp. 250 – 251.

Burt, P. J., and Adelson, E. H. [1983]. "The Laplacian Pyramid as a Compact Image Code." *IEEE Trans. Commun.*, vol. COM-31, no. 4, pp. 532 – 540.

Campbell, J. D. [1969]. "Edge Structure and the Representation of Pictures." Ph.D. dissertation, Dept. of Elec. Eng., University of Missouri, Columbia.

Candy, J. C., Franke, M. A., Haskell, B. G., and Mounts, F. W. [1971]. "Transmitting Television as Clusters of Frame-to-Frame Differences." *Bell Sys. Tech. J.*, vol. 50, pp. 1889 – 1919.

Cannon, T. M. [1974]. "Digital Image Deblurring by Non-Linear Homomorphic Filtering." Ph.D. thesis, University of Utah.

Canny, J. [1986]. "A Computational Approach for Edge Detection," *IEEE Trans. Pattern Anal.*

Machine Intell., vol. 8, no. 6, pp. 679 - 698.

Carey, W. K., Chuang, D. B., and Hamami, S. S. [1999]. "Regularity-Preserving Image Interpolation," *IEEE Trans. Image Processing*, vol. 8, no. 9, pp. 1293 - 1299.

Caselles, V., Lisani, J.-L., Morel, J.-M., and Sapiro, G. [1999]. "Shape Preserving Local Histogram Modification," *IEEE Trans. Image Processing*, vol. 8, no. 2, pp. 220 - 230.

Castleman, K. R. [1996]. *Digital Image Processing*, 2nd ed. Prentice Hall, Upper Saddle River, N.J.

Centeno, J. A. S., and Haertel, V. [1997]. "An Adaptive Image Enhancement Algorithm," *Pattern Recog.*, vol. 30, no. 7 pp. 1183 - 1189.

Chan, R. C., Karl, W. C., and Lees, R. S. [2000]. "A New Model-Based Technique for Enhanced Small-Vessell Measurements in X-Ray Cine-Angiograms," *IEEE Trans. Medical Imaging*, vol. 19, no. 3, pp. 243 - 255.

Chang, S. G., Yu, B., and Vetterli, M. [2000]. "Spatially Adaptive Wavelet Thresholding with Context Modeling for Image Denoising." *IEEE Trans. Image Processing*, vol. 9, no. 9, pp. 1522 - 1531.

Chang, S. K. [1989]. *Principles of Pictorial Information Systems Design*, Prentice Hall, Englewood Cliffs, N.J.

Chang, T., and Kuo, C.-C. J. [1993]. "Texture Analysis and Classification with Tree-Structures Wavelet Transforms." *IEEE Trans. Image Processing*, vol. 2, no. 4, pp. 429 - 441.

Chaudhuri, B. B. [1983]. "A Note on Fast Algorithms for Spatial Domain Techniques in Image Processing." *IEEE Trans. Syst. Man Cyb.*, vol. SMC-13, no. 6, pp. 1166 - 1169.

Chen, M. C., and Wilson, A. N. [2000]. "Motion-Vector Optimization of Control Grid Interpolation and Overlapped Block Motion Compensation Using Iterated Dynamic Programming," *IEEE Trans. Image Processing.*, vol. 9, no. 7, pp. 1145 - 1157.

Chen, Y.-S., and Yu, Y.-T. [1996]. "Thinning Approach for Noisy Digital Patterns," *Pattern Recog.*, vol. 29, no. 11, pp. 1847 - 1862.

Cheng, H. D., and Huijuan Xu, H. [2000]. "A Novel Fuzzy Logic Approach to Contrast Enhancement," *Pattern Recog.*, vol. 33, no. 5, pp. 809 - 819.

Cheriet, M., Said, J. N., and Suen, C. Y. [1998]. "A Recursive Thresholding Technique for Image Segmentation," *IEEE Trans. Image Processing*, vol. 7, no. 6, pp. 918 - 921.

Cheung, J., Ferris, D., and Kurz, L. [1997]. "On Classification of Multispectral Infrared Image Data," *IEEE Trans. Image Processing*, vol. 6, no. 10, pp. 1456 - 1464.

Cheung, K. K. T., and Teoh, E. K. [1999]. "Symmetry Detection by Generalized Complex (GC) Moments: A Closed-Form Solution," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, no. 5, pp. 466 - 476.

Chow, C. K., and Kaneko, T. [1972]. "Automatic Boundary Detection of the Left Ventricle from Cineangiograms." *Comp., and Biomed. Res.*, vol. 5, pp. 388 - 410.

Chu, C.-C., and Aggarwal, J. K [1993]. "The Integration of Image Segmentation Maps Using Regions and Edge Information," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, no. 12,

pp. 1241 – 1252.

CIE [1978]. *Uniform Color Spaces — Color Difference Equations — Psychometric Color Terms*, Commission Internationale de L'Eclairage, Publication No. 15, Supplement No. 2, Paris.

Clark, J. J. [1989]. "Authenticating Edges Produced by Zero-Crossing Algorithms." *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 8, pp. 830 – 831.

Clarke, R. J. [1985]. *Transform Coding of Images*, Academic Press, New York.

Cochran, W. T., Cooley, J. W., et al. [1967]. "What Is the Fast Fourier Transform?" *IEEE Trans. Audio and Electroacoustics*, vol. AU-15, no. 2, pp. 45 – 55.

Cohen, A., and Daubechies, I. [1992]. *A Stability Criterion for Biorthogonal Wavelet Bases and Their Related Subband Coding Schemes*, Technical report, AT&T Bell Laboratories.

Cohen, A., Daubechies, I., and Feauveau, J.-C. [1992]. "Biorthogonal Bases of Compactly Supported Wavelets." *Commun. Pure and Appl. Math.*, vol. 45, pp. 485 – 560.

Coifman, R. R., and Wickerhauser, M. V. [1992]. "Entropy-Based Algorithms for Best Basis Selection." *IEEE Tran. Information Theory*, vol. 38, no. 2, pp. 713 – 718.

Cooley, J. W., Lewis, P. A. W., and Welch, P. D. [1967a]. "Historical Notes on the Fast Fourier Transform." *IEEE Trans. Audio and Electroacoustics*, vol. AU-15, no. 2, pp. 76 – 79.

Cooley, J. W., Lewis, P. A. W., and Welch, P. D. [1967b]. "Application of the Fast Fourier Transform to Computation of Fourier Integrals." *IEEE Trans. Audio and Electroacoustics*, vol. AU-15, no. 2, pp. 79 – 84.

Cooley, J. W., Lewis, P. A. W., and Welch, P. D. [1969]. "The Fast Fourier Transform and Its Applications." *IEEE Trans. Educ.*, vol. E-12, no. 1. pp. 27 – 34.

Cooley, J. W., and Tukey, J. W. [1965]. "An Algorithm for the Machine Calculation of Complex Fourier Series." *Math. of Comput.*, vol. 19, pp. 297 – 301.

Cornsweet, T. N. [1970]. *Visual Perception*, Academic Press, New York.

Cortelazzo, G. M., Lucchese, L., and Monti, C. M. [1999]. "Frequency Domain Analysis of General Planar Rigid Motion with Finite Duration," *J. Opt. Soc. Amer.-A. Optics, Image Science, and Vision*, vol. 16, no. 6, pp. 1238 – 1253.

Cowart, A. E., Snyder, W. E., and Ruedger, W. H. [1983]. "The Detection of Unresolved Targets Using the Hough Transform." *Comput. Vision Graph Image Proc.*, vol. 21, pp. 222 – 238.

Croisier, A., Esteban, D., and Galand, C. [1976]. "Perfect Channel Splitting by Use of Interpolation/Decimation/Tree Decomposition Techniques." *Int. Conf. On Inform. Sciences and Systems*, Patras, Greece, pp. 443 – 446.

Cumani, A., Guiducci, A., and Grattoni, P. [1991]. "Image Description of Dynamic Scenes," *Pattern Recog.*, vol. 24, no. 7, pp. 661 – 674.

Cutrona, L. J., and Hall, W. D. [1968]. "Some Considerations in Post-Facto Blur Removal." In *Evaluation of Motion-Degraded Images*, NASA Publ. SP-193, pp. 139 – 148.

Danielson, G. C., and Lanczos, C. [1942]. "Some Improvements in Practical Fourier Analysis

and Their Application to X-Ray Scattering from Liquids." *J. Franklin Institute*, vol. 233, pp. 365 – 380, 435 – 452.

Daubechies, I. [1988]. "Orthonormal Bases of Compactly Supported Wavelets." *Commun. On Pure and Appl. Math.*, vol. 41, pp. 909 – 996.

Daubechies, I. [1990]. "The Wavelet Transform, Time-Frequency Localization and Signal Analysis." *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961 – 1005.

Daubechies, I. [1992]. *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia, Pa.

Daubechies, I. [1993]. "Orthonormal Bases of Compactly Supported Wavelets II, Variations on a Theme." *SIAM J. Mathematical Analysis*, vol. 24, no. 2, pp. 499 – 519.

Daubechies, I. [1996]. "Where Do We Go from Here? —A Personal Point of View." *Proc. IEEE*, vol. 84, no. 4, pp. 510 – 513.

Daul, C., Graebling, P., and Hirsch, E. [1998]. "From the Hough Transform to a New Approach for the Detection of and Approximation of Elliptical Arcs," *Computer Vision and Image Understanding*, vol. 72, no. 3, pp. 215 – 236.

Davis, L. S. [1982]. "Hierarchical Generalized Hough Transforms and Line-Segment Based Generalized Hough Transforms." *Pattern Recog.*, vol. 15, no. 4, pp. 277 – 285.

Davis, T. J. [1999]. "Fast Decomposition of Digital Curves into Polygons Using the Haar Transform," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, no. 8, pp. 786 – 790.

Davisson, L. D. [1972]. "Rate-Distortion Theory and Application." *Proc. IEEE*, vol. 60, pp. 800 – 808.

Delaney, A. H., and Bresler, Y. [1995]. "Multiresolution Tomographic Reconstruction Using Wavelets." *IEEE Trans. Image Processing*, vol. 4, no. 6, pp. 799 – 813.

Delp, E. J., and Mitchell, O. R. [1979]. "Image Truncation using Block Truncation Coding." *IEEE Trans. Comm.*, vol. COM-27, pp. 1335 – 1342.

Di Zenzo, S. [1986]. "A Note on the Gradient of a Multi-Image," *Computer Vision, Graphics, and Image Processing*, vol. 33, pp. 116 – 125.

Dijkstra, E. [1959]. "Note on Two Problems in Connection with Graphs." *Numerische Mathematik*, vol. 1, pp. 269 – 271.

Djeziri, S., Nouboud, F., and Plamondon, R. [1998]. "Extraction of Signatures from Check Background Based on a Filiformity Criterion," *IEEE Trans. Image Processing*, vol. 7, no. 102, pp. 1425 – 1438.

Dougherty, E. R. [1992]. *An Introduction to Morphological Image Processing*, SPIE Press, Bellingham, Wash.

Dougherty, E. R. (ed.) [2000]. *Random Processes for Image and Signal Processing*, IEEE Press, New York.

Drew, M. S., Wei, J., and Li, Z.-N. [1999]. "Illumination Invariant Image Retrieval and Video Segmentation," *Pattern Recog.*, vol. 32, no. 8, pp. 1369 – 1388.

Duda, R. O., and Hart, P. E. [1972]. "Use of the Hough Transformation to Detect Lines and

Curves in Pictures." *Comm. ACM*, vol. 15, no. 1, pp. 11 – 15.

Duda, R. O, Hart, P. E., and Stork, D. G. [2001]. *Pattern Classification*, John Wiley & Sons, New York.

Edelman, S. [1999]. *Representation and Recognition in Vision*, The MIT Press, Cambridge, Mass.

Elias, P. [1952]. "Fourier Treatment of Optical Processes." *J. Opt. Soc. Am.*, vol. 42, no. 2, pp. 127 – 134.

Elliott, D. F., and Rao, K. R. [1983]. *Fast Transforms: Algorithms and Applications*. Academic Press, New York.

Eng, H.-L., and Ma, K.-K. [2001]. "Noise Adaptive Soft-Switching Median Filter," *IEEE Trans. Image Processing*, vol. 10, no. 2, pp. 242 – 251.

Equitz, W. H. [1989]. "A New Vector Quantization Clustering Algorithm." *IEEE Trans. Acous. Speech Signal Processing*, vol. ASSP-37, no. 10, pp. 1568 – 1575.

Etienne, E. K., and Nachtegael, M. (eds.) [2000]. *Fuzzy Techniques in Image Processing*, Springer-Verlag, New York.

Falconer, D. G. [1970]. "Image Enhancement and Film Grain Noise." *Opt. Acta*, vol. 17, pp. 693 – 705.

Fairchild, M. D. [1998]. *Color Appearance Models*, Addison-Wesley, Reading, Mass.

Federal Bureau of Investigation [1993]. *WSQ Gray-Scale Fingerprint Image Compression Specification*, IAFIS-IC-0110v2, Washington, D. C.

Felsen, L. B., and Marcuvitz, N. [1994]. *Radiation and Scattering of Waves*, IEEE Press, New York.

Ferreira, A., and Ubéda, S. [1999]. "Computing the Medial Axis Transform in Parallel with Eight Scan Operators," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, no. 3, pp. 277 – 282.

Fischler, M. A. [1980]. "Fast Algorithms for Two Maximal Distance Problems with Applications to Image Analysis." *Pattern Recog.*, vol. 12, pp. 35 – 40.

Fisher, R. A. [1936]. "The Use of Multiple Measurements in Taxonomic Problems," *Ann. Eugenics*, vol. 7, Part 2, pp. 179 – 188. (Also in *Contributions to Mathematical Statistics*, John Wiley & Sons, New York, 1950.)

Fortner, B., and Meyer, T. E. [1997]. *Number by Colors*, Springer-Verlag, New York.

Fox, E. A. [1991]. "Advances in Interactive Digital Multimedia Systems." *Computer*, vol. 24, no. 10, pp. 9 – 21.

Fram, J. R., and Deutsch, E. S. [1975]. "On the Quantitative Evaluation of Edge Detection Schemes and Their Comparison with Human Performance." *IEEE Trans. Computers*, vol. C-24, no. 6, pp. 616 – 628.

Freeman, A. (translator) [1878]. J. Fourier, *The Analytical Theory of Heat*, Cambridge: University Press, London.

Freeman, C. [1987]. *Imaging Sensors and Displays*, ISBN 0 – 89252 – 800 – 1, SPIE Press, Bellingham, Wash.

Freeman, H. [1961]. "On the Encoding of Arbitrary Geometric Configurations." *IEEE Trans.*

Elec. Computers, vol. EC-10, pp. 260 – 268.

Freeman, H. [1974]. "Computer Processing of Line Drawings." *Comput. Surveys*, vol. 6, pp. 57 – 97.

Freeman, H., and Shapira, R. [1975]. "Determining the Minimum-Area Encasing Rectangle for an Arbitrary Closed Curve." *Comm. ACM*, vol. 18, no. 7, pp. 409 – 413.

Freeman, J. A., and Skapura, D. M. [1991]. *Neural Networks: Algorithms, Applications, and Programming Techniques*, Addison-Wesley, Reading, Mass.

Frei, W., and Chen, C. C. [1977]. "Fast Boundary Detection: A Generalization and a New Algorithm." *IEEE Trans. Computers*, vol. C-26, no. 10, pp. 988 – 998.

Frendendall, C. L., and Behrend, W. L. [1960]. "Picture Quality—Procedures for Evaluating Subjective Effects of Interference." *Proc. IRE*, vol. 48, pp. 1030 – 1034.

Fu, K. S. [1982]. *Syntactic Pattern Recognition and Applications*, Prentice Hall, Engle-wood Cliffs, N.J.

Fu, K. S., and Bhargava, B. K. [1973]. "Tree Systems for Syntactic Pattern Recognition." *IEEE Trans. Comput.*, vol. C-22, no. 12, pp. 1087 – 1099.

Fu, K. S., Gonzalez, R. C., and Lee, C. S. G. [1987]. *Robotics: Control, Sensing, Vision, and Intelligence*, McGraw-Hill, New York.

Fu, K. S., and Mui, J. K. [1981]. "A Survey of Image Segmentation." *Pattern Recog.*, vol. 13, no. 1, pp. 3 – 16.

Garcia, P. [1999]. "The Use of Boolean Model for Texture Analysis of Grey Images," *Computer Vision and Image Understanding*, vol. 74, no. 3, pp. 227 – 235.

Gdalyahu, Y., and Weinshall, D. [1999]. "Flexible Syntactic Matching of Curves and Its Application to Automated Hierarchical Classification of Silhouettes," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, no. 12, pp. 1312 – 1328.

Gegenfurtner, K. R., and Sharpe, L. T. (eds.) [1999]. *Color Vision: From Genes to Perception*, Cambridge University Press, New York.

Geladi, P., and Grahn, H. [1996]. *Multivariate Image Analysis*, John Wiley & Sons, New York.

Geman, D., and Reynolds, G. [1992]. "Constrained Restoration and the Recovery of Discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 3, pp. 367 – 383.

Gentleman, W. M. [1968]. "Matrix Multiplication and Fast Fourier Transformations." *Bell System Tech. J.*, vol. 47, pp. 1099 – 1103.

Gentleman, W. M., and Sande, G. [1966]. "Fast Fourier Transform for Fun and Profit." *Fall Joint Computer Conf.*, vol. 29, pp. 563 – 578, Spartan, Washington, D. C.

Gharavi, H., and Tabatabai, A. [1988]. "Sub-Band Coding of Monochrome and Color Images." *IEEE Trans. Circuits Sys.*, vol. 35, no. 2, pp. 207 – 214.

Giannakis, G. B., and Heath, R. W., Jr. [2000]. "Blind Identification of Multichannel FIR Blurs and Perfect Image Restoration," *IEEE Trans. Image Processing*, vol. 9, no. 11, pp. 1877 – 1896.

Giardina, C. R., and Dougherty, E. R. [1988]. *Morphological Methods in Image and Signal*

Processing, Prentice Hall, Upper Saddle River, N.J.

Gonzalez, R. C.[1985]. "Computer Vision." *Yearbook of Science and Technology*, McGraw-Hill, New York, pp. 128 – 132.

Gonzalez, R. C. [1985]. "Industrial Computer Vision." In *Advances in Information Systems Science*, Tou, J. T. (ed.), Plenum, New York, pp. 345 – 385.

Gonzalez, R. C. [1986]. "Image Enhancement and Restoration." In *Handbook of Pattern Recognition and Image Processing*, Young, T. Y., and Fu, K. S. (eds.), Academic Press, New York, pp. 191 – 213.

Gonzalez, R. C., Edwards, J. J., and Thomason, M. G. [1976]. "An Algorithm for the Inference of Tree Grammars." *Int. J. Comput. Info. Sci.*, vol. 5, no. 2, pp. 145 – 163.

Gonzalez, R. C., and Fittes, B. A. [1977]. "Gray-Level Transformations for Interactive Image Enhancement." *Mechanism and Machine Theory*, vol. 12, pp. 111 – 122.

Gonzalez, R. C., and Safabakhsh, R. [1982]. "Computer Vision Techniques for Industrial Applications." *Computer*, vol. 15, no. 12, pp. 17 – 32.

Gonzalez, R. C., and Thomason, M. G. [1978]. *Syntactic Pattern Recognition*: An Introduction. Addison-Wesley, Reading, Mass.

Gonzalez, R. C., and Woods, R. E. [1992]. *Digital Image Processing*. Addison-Wesley, Reading, Mass.

Good, I. J. [1958]. "The Interaction Algorithm and Practical Fourier Analysis." *J. R. Stat. Soc. (Lond.)*, vol. B20, pp. 361 – 367; *Addendum*, vol. 22, 1960, pp. 372 – 375.

Goodson, K. J., and Lewis, P. H. [1990]. "A Knowledge-Based Line Recognition System." *Pattern Recog. Letters*, vol. 11, no. 4, pp. 295 – 304.

Gordon, I. E. [1997]. *Theories of Visual Perception*, 2nd ed., John Wiley & Sons, New York.

Gori, M., and Scarselli, F. [1998]. "Are Multilayer Perceptrons Adequate for Pattern Recognition and Verification?" *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 11, pp. 1121 – 1132.

Goutsias, J., Vincent, L., and Bloomberg, D. S. (eds) [2000]. *Mathematical Morphology and Its Applications to Image and Signal Processing*, Kluwer Academic Publishers, Boston, Mass.

Graham, R. E. [1958]. "Predictive Quantizing of Television Signals." *IRE Wescon Conv. Rec.*, vol. 2, pt. 2, pp. 147 – 157.

Graham, R. L., and Yao, F. F. [1983]. "Finding the Convex Hull of a Simple Polygon." *J. Algorithms*, vol. 4, pp. 324 – 331.

Gray, R. M. [1984]. "Vector Quantization." *IEEE Trans. Acous. Speech Signal Processing*, vol. ASSP-1, no. 2, pp. 4 – 29.

Gröchenig, K., and Madych, W. R. [1992]. "Multiresolution Analysis, Haar Bases and Self-Similar Tilings of R^n." *IEEE Trans. Information Theory*, vol. 38, no. 2, pp. 556 – 568.

Grossman, A., and Morlet, J. [1984]. "Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape." *SIAM J. Appl. Math*. vol. 15, pp. 723 – 736.

Guil, N., Villalba, J., and Zapata, E. L. [1995]. "A Fast Hough Transform for Segment Detec-

tion," *IEEE Trans. Image Processing*, vol. 4, no. 11, pp. 1541 – 1548.

Guil, N., and Zapata, E. L. [1997]. "Lower Order Circle and Ellipse Hough Transform," *Pattern Recog.*, vol. 30, no. 10, pp. 1729 – 1744.

Gunn, S.R. [1998]. "Edge Detection Error in the Discrete Laplacian of a Gaussian," *Proc. 1998 Int'l Conference on Image Processing*, vol. II, pp. 515 – 519.

Gunn, S. R. [1999]. "On the Discrete Representation of the Laplacian of a Gaussian," *Pattern Recog.*, vol. 32, no. 8, pp. 1463 – 1472.

Gupta, L., Mohammad, R. S., and Tammana, R. [1990]. "A Neural Network Approach to Robust Shape Classification." *Pattern Recog.*, vol. 23, no. 6, pp. 563 – 568.

Gupta, L., and Srinath, M. D. [1988]. "Invariant Planar Shape Recognition Using Dynamic Alignment." *Pattern Recog.*, vol. 21, pp. 235 – 239.

Gupta, L., Wang., J., Charles, A., and Kisatsky, P. [1994]. "Three-Layer Perceptron Based Classifiers for the Partial Shape Classification Problem," *Pattern Recog.*, vol. 27, no. 1, pp. 91 – 97.

Haar, A. [1910]. "Zur Theorie der Orthogonalen Funktionensysteme." *Math. Annal.*, vol. 69, pp. 331 – 371.

Habibi, A. [1971]. "Comparison of Nth Order DPCM Encoder with Linear Transformations and Block Quantization Techniques." *IEEE Trans. Comm. Tech.*, vol. COM-19, no. 6, pp. 948 – 956.

Habibi, A. [1974]. "Hybrid Coding of Pictorial Data." *IEEE Trans. Comm.*, vol. COM-22, no. 5, pp. 614 – 624.

Haddon, J. F., and Boyce, J. F. [1990]. "Image Segmentation by Unifying Region and Boundary Information." *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 10, pp. 929 – 948.

Hall, E.L. [1979]. *Computer Image Processing and Recognition*, Academic Press, New York.

Hamming, R. W. [1950]. "Error Detecting and Error Correcting Codes." *Bell Sys. Tech. J.*, vol. 29, pp. 147 – 160.

Hannah, I., Patel, D., and Davies, R. [1995]. "The Use of Variance and Entropy Thresholding Methods for Image Segmentation," *Pattern Recog.*, vol. 28, no. 8, pp. 1135 – 1143.

Haralick, R. M., and Lee, J. S. J. [1990]. "Context Dependent Edge Detection and Evaluation." *Pattern Recog.*, vol. 23, no. 1 – 2, pp. 1 – 20.

Haralick, R. M., and Shapiro, L. G. [1985]. "Survey: Image Segmentation." *Comput. Vision, Graphics, Image Proc.*, vol. 29, pp. 100 – 132.

Haralick, R. M., and Shapiro, L. G. [1992]. *Computer and Robot Vision*, vols. 1 & 2, Addison-Wesley, Reading, Mass.

Haralick, R. M., Sternberg, S. R., and Zhuang, X. [1987]. "Image Analysis Using Mathematical Morphology." *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, no. 4, pp. 532 – 550.

Haralick, R. M., Shanmugan, R., and Dinstein, I. [1973]. "Textural Features for Image Clas-

sification." *IEEE Trans Syst. Man Cyb.*, vol. SMC-3, no. 6, pp. 610 – 621.

Harikumar, G., and Bresler, Y. [1999]. "Perfect Blind Restoration of Images Blurred by Multiple Filters: Theory and Efficient Algorithms," *IEEE Trans. Image Processing*, vol. 8, no. 2, pp. 202 – 219.

Harmuth, H. F. [1970]. *Transmission of Information by Orthogonal Signals*, SpringerVerlag, New York.

Haris, K., Efstratiadis, S. N, Maglaveras, N., and Katsaggelos, A. K. [1998]. "Hybrid Image Segmentation Using Watersheds and Fast Region Merging," *IEEE Trans. Image Processing*, vol. 7, no. 12, pp. 1684 – 1699.

Hart, P. E., Nilsson, N. J., and Raphael, B. [1968]. "A Formal Basis for the Heuristic Determination of Minimum-Cost Paths." *IEEE Trans. Syst. Man Cyb*, vol. SMC-4, pp. 100 – 107.

Hartenstein, H., Ruhl, M., and Saupe, D. [2000]. "Region-Based Fractal Image Compression." *IEEE Trans. Image Processing*, vol. 9, no. 7, pp. 1171 – 1184.

Haskell, B. G., and Netravali, A. N. [1997]. *Digital Pictures: Representation, Compression, and Standards*, Perseus Publishing, New York.

Haykin, S. [1996]. *Adaptive Filter Theory*, Prentice Hall, Upper Saddle River, N.J.

Healy, D. J., and Mitchell, O. R. [1981]. "Digital Video Bandwidth Compression Using Block Truncation Coding." *IEEE Trans. Comm.*, vol. COM-29, no. 12, pp. 1809 – 1817.

Heath, M. D., Sarkar, S., Sanocki, T., and Bowyer, K. W. [1997]. "A Robust Visual Method for Assessing the Relative Performance of Edge-Detection Algorithms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 12, pp. 1338 – 1359.

Heath, M., Sarkar, S., Sanoki, T., and Bowyer, K. [1998]. "Comparison of Edge Detectors: A Methodology and Initial Study," *Computer Vision and Image Understanding*, vol. 69, no. 1, pp. 38 – 54.

Hebb, D. O. [1949]. *The Organization of Behavior: A Neuropsychological Theory*, John Wiley & Sons, New York.

Heijmans, H. J. A. M., and Goutsias, J. [2000]. "Nonlinear Multiresolution Signal Decomposition Schemes—Part II: Morphological Wavelets." *IEEE Trans. Image Processing*, vol. 9, no. 11, pp. 1897 – 1913.

Highnam, R., and Brady, M. [1997]. "Model-Based Image Enhancement of Far Infrared Images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 4, pp. 410 – 415.

Hojjatoleslami, S. A., and Kittler, J. [1998]. "Region Growing: A New Approach," *IEEE Trans. Image Processing*, vol. 7, no. 7, pp. 1079 – 1084.

Hoover, R. B., and Doty, F. [1996]. *Hard X-Ray/Gamma-Ray and Neutron Optics, Sensors, and Applications*, ISBN 0-8194-2247-9, SPIE Press, Bellingham, Wash.

Horn, B. K. P. [1986]. *Robot Vision*, McGraw-Hill, New York.

Hotelling, H. [1933]. "Analysis of a Complex of Statistical Variables into Principal Components." *J. Educ. Psychol.*, vol. 24, pp. 417 – 441, 498 – 520.

Hough, P. V. C. [1962]. "Methods and Means for Recognizing Complex Patterns." U. S. Patent 3 069 654.

Hsu, C. C., and Huang, J. S. [1990]. "Partitioned Hough Transform for Ellipsoid Detection." *Pattern Recog.*, vol. 23, no. 3-4, pp. 275-282.

Hu. J., and Yan, H. [1997]. "Polygonal Approximation of Digital Curves Based on the Principles of Perceptual Organization," *Pattern Recog.*, vol. 30, no. 5, pp. 701-718.

Hu, M. K. [1962]. "Visual Pattern Recognition by Moment Invariants." *IRE Trans. Info. Theory*, vol. IT-8, pp. 179-187.

Huang, S.-C., and Sun, Y.-N. [1999]. "Polygonal Approximation Using Generic Algorithms," *Pattern Recog.*, vol. 32, no. 8, pp. 1409-1420.

Huang, T. S. [1965]. "PCM Picture Transmission." *IEEE Spectrum*, vol. 2, no. 12, pp. 57-63.

Huang, T. S. [1966]. "Digital Picture Coding." *Proc. Natl. Electron. Conf.*, pp. 793-797.

Huang, T. S., ed. [1975]. *Picture Processing and Digital Filtering*, Springer-Verlag, New York.

Huang, T. S. [1981]. *Image Sequence Analysis*, Springer-Verlag, New York.

Huang, T. S., and Hussian, A. B. S. [1972]. "Facsimile Coding by Skipping White." *IEEE Trans. Comm.*, vol. COM-23, no. 12, pp. 1452-1466.

Huang, T. S., and Tretiak, O. J. (eds.). [1972]. *Picture Bandwidth Compression*, Gordon and Breech, New York.

Huang, T. S., Yang, G. T., and Tang, G. Y. [1979]. "A Fast Two-Dimensional Median Filtering Algorithm." *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. ASSP-27, pp. 13-18.

Huang, Y., and Schultheiss, P. M. [1963]. "Block Quantization of Correlated Gaussian Random Variables." *IEEE Trans. Commun. Syst.*, vol. CS-11, pp. 289-296.

Hubbard, B. B. [1998]. *The World According to Wavelets-The Story of a Mathematical Technique in the Making*, 2nd ed, A. K. Peters, Ltd., Wellesley, Mass.

Hubel, D. H. [1988]. *Eye, Brain, and Vision*, Scientific Amer. Library, W. H. Freeman, New York.

Huertas, A., and Medione, G. [1986]. "Detection of Intensity Changes with Subpixel Accuracy using Laplacian-Gaussian Masks," *IEEE Trans. Pattern. Anal. Machine Intell.*, vol. PAMI-8, no. 5, pp. 651-664.

Huffman, D. A. [1952]. "A Method for the Construction of Minimum Redundancy Codes." *Proc. IRE*, vol. 40, no. 10, pp. 1098-1101.

Hufnagel, R. E., and Stanley, N. R. [1964]. "Modulation Transfer Function Associated with Image Transmission Through Turbulent Media," *J. Opt. Soc. Amer.*, vol. 54, pp. 52-61.

Hummel, R. A. [1974]. "Histogram Modification Techniques." Technical Report TR-329. F-44620-72C-0062, Computer Science Center, University of Maryland, College Park, Md.

Hunt, B. R. [1971]. "A Matrix Theory Proof of the Discrete Convolution Theorem." *IEEE Trans. Audio and Electroacoust.*, vol. AU-19, no. 4, pp. 285-288.

Hunt, B. R. [1973]. "The Application of Constrained Least Squares Estimation to Image Restoration by Digital Computer." *IEEE Trans. Comput.*, vol. C-22, no. 9, pp. 805 – 812.

Hunter, R., and Robinson, A. H. [1980]. "International Digital Facsimile Coding Standards." *Proc. IEEE*, vol. 68, no. 7, pp. 854 – 867.

Hurn, M., and Jennison, C. [1996]. "An Extension of Geman and Reynolds' Approach to Constrained Restoration and the Recovery of Discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, no. 6, pp. 657 – 662.

Hwang, H., and Haddad, R. A. [1995]. "Adaptive Median Filters: New Algorithms and Results," *IEEE Trans. Image Processing*, vol. 4, no. 4, pp. 499 – 502.

IEEE Computer [1974]. Special issue on digital image processing. vol. 7, no. 5.

IEEE Computer [1988]. Special issue on artificial neural systems. vol. 21, no. 3.

IEEE Trans. Circuits and Syst. [1975]. Special issue on digital filtering and image processing, vol. CAS-2, pp. 161 – 304.

IEEE Trans. Computers [1972]. Special issue on two-dimensional signal processing, vol. C-21, no. 7.

IEEE Trans. Comm. [1981]. Special issue on picture communication systems, vol. COM-29, no. 12.

IEEE Trans. on Image Processing [1994]. Special issue on image sequence compression, vol. 3, no. 5.

IEEE Trans. on Image Processing [1996]. Special issue on vector quantization, vol. 5, no. 2.

IEEE Trans. Image Processing [1996]. Special issue on nonlinear image processing, vol. 5, no. 6.

IEEE Trans. Image Processing [1997]. Special issue on automatic target detection, vol. 6, no. 1.

IEEE Trans. Image Processing [1997]. Special issue on color imaging, vol. 6, no. 7.

IEEE Trans. Image Processing [1998]. Special issue on applications of neural networks to image processing, vol. 7, no. 8.

IEEE Trans. Information Theory [1992]. Special issue on wavelet transforms and mulitresolution signal analysis, vol. 11, no. 2, Part II.

IEEE Trans. Pattern Analysis and Machine Intelligence [1989]. Special issue on mnltiresolution processing, vol. 11, no. 7.

IEEE Trans. Signal Processing [1993]. Special issue on wavelets and signal processing, vol. 41, no. 12.

IES Lighting Handbook, 9th ed. [2000]. Illuminating Engineering Society Press, New York.

ISO/IEC [1999]. *ISO/IEC 14495-1: 1999: Information technology—Lossless and nearlossless compression of continuous-tone still images: Baseline.*

ISO/IEC JTC 1/SC 29/WG 1 [2000]. *ISO/IEC FCD 15444-1: Information technology— JPEG 2000 image coding system: Core coding system.*

Jahne, B. [1997]. *Digital Image Processing: Concepts, Algorithms, and Scientific Applications*, Springer-Verlag, New York.

Jain, A. K. [1981]. "Image Data Compression: A Review." *Proc. IEEE*, vol. 69, pp. 349-389.

Jain, A. K. [1989]. *Fundamentals of Digital Image Processing*, Prentice Hall, Englewood Cliffs, N.J.

Jain, A. K., Duin, R. P. W., and Mao, J. [2000]. "Statistical Pattern Recognition: A Review," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 1, pp. 4-37.

Jain, J. R., and Jain, A. K. [1981]. "Displacement Measurement and Its Application in Interframe Image Coding." *IEEE Trans. Comm.*, vol. COM-29, pp. 1799-1808.

Jain, R. [1981]. "Dynamic Scene Analysis Using Pixel-Based Processes." *Computer*, vol. 14, no. 8, pp. 12-18.

Jain, R., Kasturi, R., and Schunk, B. [1995]. *Computer Vision*, McGraw-Hill, New York.

Jang, B. K., and Chin, R. T. [1990]. "Analysis of Thinning Algorithms Using Mathematical Morphology." *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 6, pp. 541-551.

Jayant, N. S. (ed.) [1976]. *Waveform Quantization and Coding*, IEEE Press, New York.

Jones, R., and Svalbe, I. [1994]. "Algorithms for the Decomposition of Gray-Scale Morphological Operations," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 6, pp. 581-588.

Jonk, A., van den Boomgaard, S., and Smeulders, A. [1999]. "Grammatical Inference of Dashed Lines," *Computer Vision and Image Understanding*, vol. 74, no. 3, pp. 212-226.

Kahaner, D. K. [1970]. "Matrix Description of the Fast Fourier Transform." *IEEE Trans. Audio Electroacoustics*, vol. AU-18, no. 4, pp. 442-450.

Kak, A. C., and Slaney, M. [2001]. *Principles of Computerized Tomographic Imaging*, Published by the Society for Industrial and Applied Mathematics, Philadelphia, Pa.

Karhunen, K. [1947]. "Über Lineare Methoden in der Wahrscheinlichkeitsrechnung." *Ann. Acad. Sci. Fennicae*, Ser. A137. (Translated by I. Selin in "On Linear Methods in Probability Theory." T-131, 1960, The RAND Corp., Santa Monica, Calif.)

Kasson, J., and Plouffe, W. [1992]. "An Analysis of Selected Computer Interchange Color Spaces." *ACM Trans. on Graphics*, vol. 11, no. 4, pp. 373-405.

Katzir, N., Lindenbaum, M., and Porat, M. [1994]. "Curve Segmentation Under Partial Occlusion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 5, pp. 513-519.

Khanna, T. [1990]. *Foundations of Neural Networks*, Addison-Wesley, Reading, Mass.

Kim, J. K., Park, J. M., Song, K. S., and Park, H. W. [1997]. "Adaptive Mammographic Image Enhancement Using First Derivative and Local Statistics," *IEEE Trans. Medical Imaging*, vol. 16, no. 5, pp. 495-502.

Kimme, C., Ballard, D. H., and Sklansky, J. [1975]. "Finding Circles by an Array of Accumulators." *Comm. ACM*, vol. 18, no. 2, pp. 120-122.

Kirsch, R. [1971]. "Computer Determination of the Constituent Structure of Biological Images." *Comput. Biomed. Res.*, vol. 4, pp. 315-328.

Kiver, M. S. [1965]. Color Television Fundamentals, McGraw-Hill, New York.

Klinger, A. [1976]. "Experiments in Picture Representation Using Regular Decomposition." *Com-*

put. *Graphics Image Proc*., vol. 5, pp. 68 – 105.

Knowlton, K. [1980]. "Progressive Transmission of Gray-Scale and Binary Pictures by Simple, Efficient, and Lossless Encoding Schemes." *Proc*. *IEEE*, vol. 68, no. 7, pp. 885 – 896.

Kohler, R. J., and Howell, H. K. [1963]. "Photographic Image Enhancement by Superposition of Multiple Images." *Photogr*. *Sci*. *Eng*., vol. 7, no. 4, pp. 241 – 245.

Kokaram, A. [1998]. *Motion Picture Restoration*, Springer-Verlag, New York.

Kramer, H. P., and Mathews, M. V. [1956]. "A Linear Coding for Transmitting a Set of Correlated Signals." *IRE Trans*. *Info*. *Theory*, vol. IT-2, pp. 41 – 46.

Langdon, G. C., and Rissanen, J. J. [1981]. "Compression of Black—White Images with Arithmetic Coding." *IEEE Trans*. *Comm*., vol. COM-29, no. 6, pp. 858 – 867.

Lantuéjoul, C. [1980]. "Skeletonization in Quantitative Metallography." In *Issues of Digital Image Processing*, Haralick, R. M., and Simon, J. C. (eds.), Sijthoff and Noordhoff, Groningen, The Netherlands.

Latecki, L. J., and Lakämper, R. [1999]. "Convexity Rule for Shape Decomposition Based on Discrete Contour Evolution," *Computer Vision and Image Understanding*, vol. 73, no. 3, pp. 441 – 454.

Le Gall, D., and Tabatabai, A. [1988]. "Sub-band Coding of Digital Images Using Symmetric Short Kernel Filters and Arithmetic Coding Techniques." *IEEE International Conference on Acoustics*, *Speech*, *and Signal Processing*, New York, NY, pp. 761 – 765.

Ledley, R. S. [1964]. "High-Speed Automatic Analysis of Biomedical Pictures." *Science*, vol. 146, no. 3461, pp. 216 – 223.

Lee, J.-S., Sun, Y.-N., and Chen, C.-H. [1995]. "Multiscale Corner Detection by Using Wavelet Transforms." *IEEE Trans*. *Image Processing*, vol. 4, no. 1, pp. 100 – 104.

Lee, S. U., Chung, S. Y., and Park, R. H. [1990]. "A Comparative Performance Study of Several Global Thresholding Techniques for Segmentation." *Comput*. *Vision*, *Graphics*, *Image Proc*., vol. 52, no. 2, pp. 171 – 190.

Lehmann, T. M., Gönner, C., and Spitzer, K. [1999]. "Survey: Interpolation Methods in Medical Image Processing," *IEEE Trans*. *Medical Imaging*, vol. 18, no. 11, pp. 1049 – 1076.

Lema, M. D., and Mitchell, O. R. [1984]. "Absolute Moment Block Truncation Coding and Its Application to Color Images." *IEEE Trans*. *Comm*., vol. COM-32, no. 10. pp. 1148 – 1157.

Levine, M. D. [1985]. *Vision in Man and Machine*, McGraw-Hill, New York.

Liang, K.-C., and Kuo, C.-C. J. [1991]. "Waveguide: A Joint Wavelet-Based Image Representation and Description System." *IEEE Trans*. *Image Processing*, vol. 8, no. 11, pp. 1619 – 1629.

Liang, Q., Wendelhag, J. W., and Gustavsson, T. [2000]. "A Multiscale Dynamic Programming Procedure for Boundary Detection in Ultrasonic Artery Images," *IEEE Trans*. *Medical Imaging*, vol. 19, no. 2, pp. 127 – 142.

Lillesand, T. M., and Kiefer, R. W. [1999]. *Remote Sensing and Image Interpretation*, John Wiley & Sons, New York.

Lim, J. S. [1990]. *Two-Dimensional Signal and Image Processing*, Prentice Hall, Upper Saddle River, N.J.

Limb, J. O., and Rubinstein, C. B. [1978]. "On the Design of Quantizers for DPCM Coders: A Functional Relationship Between Visibility, Probability, and Masking." *IEEE Trans. Comm.*, vol. COM-26, pp. 573 – 578.

Lindblad, T., and Kinser, J. M. [1998]. *Image Processing Using Pulse-Coupled Neural Networks*, Springer-Verlag, New York.

Linde, Y., Buzo, A., and Gray, R. M. [1980]. "An Algorithm for Vector Quantizer Design." *IEEE Trans. Comm.*, vol. COM-28, no. 1, pp. 84 – 95.

Lippmann, R. P. [1987]. "An Introduction to Computing with Neural Nets." *IEEE ASSP Magazine*, vol. 4, pp. 4 – 22.

Liu, J., and Yang, Y.-H. [1994]. "Multiresolution Color Image Segmentation," *IEEE Trans Pattern Anal. Machine Intell.*, vol. 16, no. 7, pp. 689 – 700.

Liu-Yu, S., and Antipolis, M. [1993]. "Description of Object Shapes by Apparent Boundary and Convex Hull," *Pattern Recog.*, vol. 26, no. 1, pp. 95 – 107.

Lo, R.-C., and Tsai, W.-H. [1995]. "Gray-Scale Hough Transform for Thick Line Detection in Gray-Scale Images," *Pattern Recog.*, vol. 28, no. 5, pp. 647 – 661.

Loncaric, S. [1998]. "A Survey of Shape Analysis Techniques," *Pattern Recog.*, vol. 31, no. 8, pp. 983 – 1010.

Lu, H. E., and Wang, P. S. P. [1986]. "A Comment on 'A Fast Parallel Algorithm for Thinning Digital Patterns.' " *Comm. ACM*, vol. 29, no. 3, pp. 239 – 242.

Lu, N. [1997]. *Fractal Imaging*, Academic Press, New York.

Lu, W.-S and Antoniou, A. [1992]. "Two-Dimensional Digital Filters," Marcel Dekker, New York.

MacAdam, D. L. [1942]. "Visual Sensitivities to Color Differences in Daylight," *J. Opt. Soc. Am.*, vol. 32, pp. 247 – 274.

MacAdam, D. P. [1970]. "Digital Image Restoration by Constrained Deconvolution." *J. Opt. Soc. Am.*, vol. 60, pp. 1617 – 1627.

Maki, A., Nordlund, P., and Eklundh, J.-O. [2000]. "Attentional Scene Segmentation: Integrating Depth and Motion," *Computer Vision and Image Understanding*, vol. 78, no. 3, pp. 351 – 373.

Malacara, D. [2001]. *Color Vision and Colorimetry: Theory and Applications*, SPIE Optical Engineering Press, Bellingham, Wash.

Mallat, S. [1987]. "A Compact Multiresolution Representation: The Wavelet Model." Proc. *IEEE Computer Society Workshop on Computer Vision*, IEEE Computer Society Press, Washington, D.C., pp. 2 – 7.

Mallat, S. [1989a]. "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-11, pp. 674 – 693.

Mallat, S. [1989b]. "Multiresolution Approximation and Wavelet Orthonormal Bases of L^2."

Trans. American Mathematical Society, vol. 315, pp. 69 – 87.

Mallat, S. [1989c]. "Multifrequency Channel Decomposition of Images and Wavelet models." *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 37, pp. 2091 – 2110.

Mallat, S. [1998]. *A Wavelet Tour of Signal Processing*, Academic Press, Boston, Mass.

Mallat, S. [1999]. *A Wavelet Tour of Signal Processing*. 2nd ed., Academic Press, San Diego, Calif.

Mallot, A. H. [2000]. *Computational Vision*, The MIT Press, Cambridge, Mass.

Mamistvalov, A. [1998]. "n-Dimensional Moment Invariants and Conceptual Mathematical Theory of Recognition [of] n-Dimensional Solids," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 8, pp. 819 – 831.

Maragos, P. [1987]. "Tutorial on Advances in Morphological Image Processing and Analysis." Optical Engineering, vol. 26, no. 7, pp. 623 – 632.

Marchand-Maillet, S., and Sharaiha, Y. M. [2000]. *Binary Digital Image Processing: A Discrete Approach*, Academic Press, New York.

Maren, A. J., Harston, C. T., and Pap, R. M. [1990]. *Handbook of Neural Computing Applications*, Academic Press, New York.

Marr, D. [1982]. Vision, Freeman, San Francisco.

Marr, D., and Hildreth, E. [1980]. "Theory of Edge Detection." *Proc. R. Soc. Lond.*, vol. B207, pp. 187 – 217.

Martelli, A. [1972]. "Edge Detection Using Heuristic Search Methods." *Comput. Graphics Image Proc.*, vol. 1, pp. 169 – 182.

Martelli, A. [1976]. "An Application of Heuristic Search Methods to Edge and Contour Detection." *Comm. ACM*, vol. 19, no. 2, pp. 73 – 83.

Martin, M. B., and Bell, A. E. [2001]. "New image compression techniques using multiwavelets and multiwavelet packets," *IEEE Trans. on Image Proc.*, vol. 10, no. 4, pp. 500 – 510.

Mather, P. M. [1999]. *Computer Processing of Remotely Sensed Images: An Introduction*, John Wiley & Sons, New York.

Max, J. [1960]. "Quantizing for Minimum Distortion." *IRE Trans. Info. Theory*, vol. IT-6, pp. 7 – 12.

McClelland, J. L., and Rumelhart, D. E. (eds.) [1986]. *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, vol. 2: *Psychological and Biological Models*, The MIT Press, Cambridge, Mass.

McCulloch, W. S., and Pitts, W. H. [1943]. "A Logical Calculus of the Ideas Imminent in Nervous Activity." *Bulletin of Mathematical Biophysics*, vol. 5, pp. 115 – 133.

McFarlane, M. D. [1972]. "Digital Pictures Fifty Years Ago." *Proc. IEEE*, vol. 60, no. 7, pp. 768 – 770.

McGlamery, B. L. [1967]. "Restoration of Turbulence-Degraded Images." *J. Opt. Soc. Am.*, vol. 57, no. 3, pp. 293 – 297.

Meijering, E. H. W., Niessen, W. J., and Viergever, M. A. [1999]. "Retrospective Motion

Correction in Digital Subtraction Angiography: A Review," *IEEE Trans. Medical Imaging*, vol. 18, no. 1, pp. 2-21. [chap 3]

Memon, N., Neuhoff, D. L., and Shende, S. [2000]. "An Analysis of Some Common Scanning Techniques for Lossless Image Coding," *IEEE Trans. Image Processing*, vol. 9, no. 11, pp. 1837-1848.

Mesarovic', V. Z. [2000]. "Iterative Linear Minimum Mean-Square-Error Image Restoration from Partially Known Blur," *J. Opt. Soc. Amer.-A. Optics, Image Science, and Vision*, vol. 17, no. 4, pp. 711-723.

Meyer, Y. [1987]. "L'analyses par Ondelettes." *Pour la Science*.

Meyer, Y. [1990]. *Ondelettes et ope' rateurs*, Hermann, Paris.

Meyer, Y. (ed.) [1992a]. *Wavelets and Applications: Proceedings of the International Conference*, Marseille, France, Mason, Paris, and Springer-Verlag, Berlin.

Meyer, Y. (translated by D. H. Salinger) [1992b]. Wavelets and Operators, Cambridge University Press, Cambridge, UK.

Meyer, Y. (translated by R. D. Ryan) [1993]. *Wavelets: Algorithms and Applications*, Society for Industrial and Applied Mathematics, Philadelphia.

Meyer, F. G., Averbuch, A. Z., and Strömberg, J.-O. [2000]. "Fast Adaptive Wavelet Pack-et Image Compression," *IEEE Trans. Image Processing*, vol. 9, no. 7, pp. 792-800.

Meyer, F., and Beucher, S. [1990]. "Morphological Segmentation." *J. Visual Comm., and Image Representation*, vol. 1, no. 1, pp. 21-46.

Meyer, H., Rosdolsky, H. G., and Huang, T. S. [1973]. "Optimum Run Length Codes." *IEEE Trans. Comm.*, vol. COM-22, no. 6, pp. 826-835.

Minsky, M., and Papert, S. [1969]. *Perceptrons: An Introduction to Computational Geometry*, the MIT Press, Cambridge, Mass.

Mirmehdi, M., and Petrou, M. [2000]. "Segmentation of Color Textures," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 2, pp. 142-159.

Misiti, M., Misiti, Y., Oppenheim, G., and Poggi, J.-M. [1996]. *Wavelet Toolbox User's Guide*, The MathWorks, Inc., Natick, Mass.

Mitiche, A. [1994]. *Computational Analysis of Visual Motion*, Perseus Publishing, New York.

Mitra, S. K., and Sicuranza, G. L. (eds.) [2000]. *Nonlinear Image Processing*, Academic Press, New York.

Murase, H., and Nayar, S. K. [1994]. "Illumination Planning for Object Recognition Using Parametric Eigen Spaces," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 12, pp. 1219-1227.

Murino, V., Ottonello, C., and Pagnan, S. [1998]. "Noisy Texture Classfication: A Higher-Order Statistical Approach," *Pattern Recog.*, vol. 31, no. 4, pp. 383-393.

Nagao, M., and Matsuyama, T. [1980]. *A Structural Analysis of Complex Aerial Photographs*, Plenum Press, New York.

Najman, L., and Schmitt, M. [1996]. "Geodesic Saliency of Watershed Contours and Hierarchi-

cal Segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, no. 12, pp. 1163–1173.

Narendra, P. M., and Fitch, R. C. [1981]. "Real-Time Adaptive Contrast Enhancement." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-3, no. 6, pp. 655–661.

Netravali, A. N. [1977]. "On Quantizers for DPCM Coding of Picture Signals." *IEEE Trans. Info. Theory*, vol. IT-23, no. 3, pp. 360–370.

Netravali, A. N., and Limb, J. O. [1980]. "Picture Coding: A Review." *Proc. IEEE*, vol. 68, no. 3, pp. 366–406.

Nevatia, R. [1982]. *Machine Perception*, Prentice Hall, Upper Saddle River, N.J.

Ngan, K. N., Meier, T., and Chai, D. [1999]. *Advanced Video Coding: Principles and Techniques*. Elsevier, Boston.

Nilsson, N. J. [1965]. *Learning Machines: Foundations of Trainable Pattern-Classifying Systems*, McGraw-Hill, New York.

Nilsson, N. J. [1971]. *Problem Solving Methods in Artificial Intelligence*, McGraw-Hill, New York.

Nilsson, N. J. [1980]. *Principles of Artificial Intelligence*, Tioga, Palo Alto, Calif.

Noble, B., and Daniel, J. W. [1988]. *Applied Linear Algebra*, 3rd ed., Prentice Hall, Upper Saddle River, N.J.

Odegard, J. E., Gopinath, R. A., and Burrus, C. S. [1992]. "Optimal Wavelets for Signal Decomposition and the Existence of Scale-Limited Signals." *Proceedings of IEEE Int. Conf. On Signal Proc.*, ICASSP-92, San Francisco, CA, vol. IV, 597–600.

Olson, C. F. [1999]. "Constrained Hough Transforms for Curve Detection," *Computer Vision and Image Understanding*, vol. 73, no. 3, pp. 329–345.

O'Neil, J. B. [1971]. "Entropy Coding in Speech and Television Differential PCM Systems." *IEEE Trans. Info. Theory*, vol. IT-17, pp. 758–761.

Oommen, R. J., and Loke, R. K. S. [1997]. "Pattern Recognition of Strings with Substitutions, Insertions, Deletions, and Generalized Transpositions," *Pattern Recog.*, vol. 30, no. 5, pp. 789–800.

Oppenheim, A. V., and Schafer, R. W. [1975]. *Digital Signal Processing*, Prentice Hall, Englewood Cliffs, N.J.

Oppenheim, A. V., Schafer, R. W., and Stockham, T. G., Jr. [1968]. "Nonlinear Filtering of Multiplied and Convolved Signals." *Proc. IEEE*, vol. 56, no. 8, pp. 1264–1291.

Oyster, C. W. [1999]. *The Human Eye: Structure and Function*, Sinauer Associates, Sunderland, Mass.

Paez, M. D., and Glisson, T. H. [1972]. "Minimum Mean-Square-Error Quantization in Speech PCM and DPCM Systems." *IEEE Trans. Comm.*, vol. COM-20, pp. 225–230.

Pao, Y. H. [1989]. *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley, Reading, Mass.

Papamarkos, N., and Atsalakis, A. [2000]. "Gray-Level Reduction Using Local Spatial Fea-

tures," *Computer Vision and Image Understanding*, vol. 78, no. 3, pp. 336 – 350.

Papoulis, A. [1991]. *Probability, Random Variables, and Stochastic Processes*, 3rd ed., McGraw-Hill, New York.

Park, H., and Chin, R. T. [1995]. "Decomposition of Arbitrarily-Shaped Morphological Structuring," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, no. 1, pp. 2 – 15.

Parker, J. R. [1991]. "Gray Level Thresholding in Badly Illuminated Images." *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 8, pp. 813 – 819.

Pattern Recognition [2000]. Special issue on Mathematical Morphology and Nonlinear Image Processing, vol. 33, no. 6, pp. 875 – 1117.

Pavlidis, T. [1977]. *Structural Pattern Recognition*, Springer-Verlag, New York.

Pavlidis, T. [1982]. *Algorithms for Graphics and Image Processing*, Computer Science Press, Rockville, Md.

Pavlidis, T., and Liow, Y. T. [1990]. "Integrating Region Growing and Edge Detection." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 3, pp. 225 – 233.

Peebles, P. Z. [1993]. *Probability, Random Variables, and Random Signal Principles*, 3rd ed., McGraw-Hill, New York.

Pennebaker, W. B., and Mitchell, J. L. [1992]. *JPEG: Still Image Data Compression Standard*. Van Nostrand Reinhold, New York.

Pennebaker, W. B., Mitchell, J. L., Langdon, G. G., Jr., and Arps, R. B. [1988]. "An Overview of the Basic Principles of the Q-coder Adaptive Binary Arithmetic Coder." *IBM J. Res. Dev.*, vol. 32, no. 6, pp. 717 – 726.

Perez, A., and Gonzalez, R. C. [1987]. "An Iterative Thresholding Algorithm for Image Segmentation." *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, no. 6, pp. 742 – 751.

Perona, P., and Malik, J. [1990]. "Scale-Space and Edge Detection Using Anisotropic Diffusion." *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 7, pp. 629 – 639.

Persoon, E., and Fu, K. S. [1977]. "Shape Discrimination Using Fourier Descriptors." *IEEE Trans. Systems Man Cyb.*, vol. SMC-7, no. 2, pp. 170 – 179.

Petrou, M., and Bosdogianni, P. [1999]. *Image Processing: The Fundamentals*, John Wiley & Sons, UK.

Petrou, M., and Kittler, J. [1991]. "Optimal Edge Detector for Ramp Edges." *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 5, pp. 483 – 491.

Piech, M. A. [1990]. "Decomposing the Laplacian." *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 8, pp. 830 – 831.

Pitas, I., and Vanetsanopoulos, A. N. [1990]. *Nonlinear Digital Filters: Principles and Applications*, Kluwer Academic Publishers, Boston, Mass.

Plataniotis, K. N, and Venetsanopoulos, A. N. [2000]. *Color Image Processing and Applications*, Springer-Verlag, New York.

Pokorny, C. K., and Gerald, C. F. [1989]. *Computer Graphics: The Principles Behind the Art*

and Science, Franklin, Beedle & Associates, Irvine, Calif.

Poynton, C. A. [1996]. *A Technical Introduction to Digital Video*, John Wiley & Sons, New York.

Prasad, L., and Iyengar, S. S. [1997]. *Wavelet Analysis with Applications to Image Processing*, CRC Press, Boca Raton, Fla.

Pratt, W. K. [1991]. *Digital Image Processing*, 2nd ed., John Wiley & Sons, New York.

Pratt, W. K. [2001]. *Digital Image Processing*, 3rd ed., John Wiley & Sons, New York.

Preparata, F. P., and Shamos, M. I. [1985]. *Computational Geometry: An Introduction*, Springer-Verlag, New York.

Preston, K. [1983]. "Cellular Logic Computers for Pattern Recognition." *Computer*, vol. 16, no. 1, pp. 36–47.

Prewitt, J. M. S. [1970]. "Object Enhancement and Extraction." in *Picture Processing and Psychopictorics*, Lipkin, B. S., and Rosenfeld, A. (eds.), Academic Press, New York.

Principe, J. C., Euliano, N. R., and Lefebre, W. C. [1999]. *Neural and Adaptive Systems: Fun-damentals through Simulations*, John Wiley & Sons, New York.

Pritchard, D. H. [1977]. "U. S. Color Television Fundamentals—A Review." *IEEE Trans. Consumer Electronics*, vol. CE-23, no. 4, pp. 467–478.

Proc. IEEE [1967]. Special issue on redundancy reduction, vol. 55, no. 3.

Proc. IEEE [1972]. Special issue on digital picture processing, vol. 60, no. 7.

Proc. IEEE [1980]. Special issue on the encoding of graphics, vol. 68, no. 7.

Proc. IEEE [1985]. Special issue on visual communication systems, vol. 73, no. 2.

Qian, R. J., and Huang, T. S. [1996]. "Optimal Edge Detection in Two-Dimensional Images," *IEEE Trans. Image Processing*, vol. 5, no. 7, pp. 1215–1220.

Rabbani, M., and Jones, P. W. [1991]. *Digital Image Compression Techniques*, SPIE Press, Bellingham, Wash.

Rajala, S. A., Riddle, A. N., and Snyder, W. E. [1983]. "Application of One-Dimensional Fourier Transform for Tracking Moving Objects in Noisy environments." *Comp., Vision, Image Proc.*, vol. 21, pp. 280–293.

Reddy, B. S., and Chatterji, B. N. [1996]. "An FFT-Based Technique for Translation, Rotation, and Scale Invariant Image Registration," *IEEE Trans. Image Processing*, vol. 5, no. 8, pp. 1266–1271.

Regan, D. D. [2000]. *Human Perception of Objects: Early Visual Processing of Spatial Form Defined by Luminance, Color, Texture, Motion, and Binocular Disparity*, Sinauer Associates, Sunderland, Mass.

Ritter, G. X., and Wilson, J. N. [2001]. *Handbook of Computer Vision Algorithms in Image Algebra*, CRC Press, Boca Raton, Fla.

Roberts, L. G. [1965]. "Machine Perception of Three-Dimensional Solids." In *Optical and Electro-Optical Information Processing*, Tippet, J. T. (ed.), MIT Press, Cambridge, Mass.

Robertson, A. R. [1977]. "The CIE 1976 Color Difference Formulae," *Color Res. Appl.*,

vol. 2, pp. 7–11.

Robinson, G. S. [1976]. "Detection and Coding of Edges Using Directional Masks." University of Southern California, Image Processing Institute, Report no. 660.

Robinson, J. A. [1965]. "A Machine-Oriented Logic Based on the Resolution Principle." *J. ACM*, vol. 12, no. 1, pp. 23–41.

Rock, I. [1984]. *Perception*, W. H. Freeman, New York.

Roese, J. A., Pratt, W. K., and Robinson, G. S. [1977]. "Interframe Cosine Transform Image Coding." *IEEE Trans. Comm.*, vol. COM-25, pp. 1329–1339.

Rosenblatt, F. [1959]. "Two Theorems of Statistical Separability in the Perceptron." In *Mechanisation of Thought Processes: Proc. of Symposium No. 10*, held at the National Physical Laboratory, November 1958, H. M. Stationery Office, London, vol. 1, pp. 421–456.

Rosenblatt, F. [1962]. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*, Spartan, Washington, D. C.

Rosenfeld, A. (ed.) [1984]. *Multiresolution Image Processing and Analysis*, Springer-Verlag, New York.

Rosenfeld, A. [1999]. "Image Analysis and Computer Vision: 1998." *Computer Vision and Image Understanding*, vol. 74, no. 1, pp. 36–95.

Rosenfeld, A. [2000]. "Image Analysis and Computer Vision: 1999." *Computer Vision and Image Understanding*, vol. 78, no. 2, pp. 222–302.

Rosenfeld, A., and Kak, A. C. [1982]. *Digital Picture Processing*, vols. 1 and 2, 2nd ed., Academic Press, New York.

Rosin, P. L. [1997]. "Techniques for Assessing Polygonal Approximations of Curves," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 6, pp. 659–666.

Rudnick, P. [1966]. "Note on the Calculation of Fourier Series." *Math. Comput.*, vol. 20, pp. 429–430.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. [1986]. "Learning Internal Representations by Error Propagation." In *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, vol. 1: *Foundations*, Rumelhart, D. E., et al. (eds.), MIT Press, Cambridge, MA., pp. 318–362.

Rumelhart, D. E., and McClelland, J. L. (eds.) [1986]. *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, vol. 1: *Foundations*, MIT Press, Cambridge, Mass.

Runge, C. [1903]. *Zeit. für Math., and Physik*, vol. 48, pp. 433.

Runge, C. [1905]. *Zeit. für Math., and Physik*, vol. 53, pp. 117.

Runge, C., and König, H. [1924]. "Die Grundlehren der Mathematischen Wissenschaften." *Vorlesungen über Numerisches Rechnen*, vol. 11, Julius Springer, Berlin.

Russ, J. C. [1999]. *The Image Processing Handbook*, 3rd ed., CRC Press, Boca Raton, Fla.

Sahni, S., and Jenq, J.-F. [1992]. "Serial and Parallel Algorithms for the Medial Axis Transform," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 12, pp. 1218–1224.

Sahoo, P. K., Soltani, S., Wong, A. K. C., and Chan, Y. C. [1988]. "A Survey of Thresholding Techniques." *Comput. Vision, Graphics, Image Proc.*, vol. 4, pp. 233 – 260.

Saito, N., and Cunningham, M. A. [1990]. "Generalized E-Filter and its Application to Edge Detection." *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 8, pp. 814 – 817.

Sakrison, D. J., and Algazi, V. R. [1971]. "Comparison of Line-by-Line and Two-Dimensional Encoding of Random Images." *IEEE Trans. Info. Theory*, vol. IT-17, no. 4, pp. 386 – 398.

Salari, E., and Siy, P. [1984]. "The Ridge-Seeking Method for Obtaining the Skeleton of Digital Images." *IEEE Trans. Syst. Man Cyb.*, vol. SMC-14, no. 3, pp. 524 – 528.

Salinas, R. A., Abidi, M. A., and Gonzalez, R. C. [1996]. "Data Fusion: Color Edge Detection and Surface Reconstruction Through Regularization," *IEEE Trans. Industrial Electronics*, vol. 43, no. 3, pp. 355 – 363, 1996.

Sato, Y. [1992]. "Piecewise Linear Approximation of Plane Curves by Perimeter Optimization," *Pattern Recog.*, vol. 25, no. 12, pp. 1535 – 1543.

Sauvola, J., and Pietikainen, M. [2000]. "Adaptive Document Image Binarization," *Pattern Recog.*, vol. 33, no. 2, pp. 225 – 236.

Schalkoff, R. J. [1989]. *Digital Image Processing and Computer Vision*, John Wiley & Sons, New York.

Schonfeld, D., and Goutsias, J. [1991]. "Optimal Morphological Pattern Restoration from Noisy Binary Images." *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 1, pp. 14 – 29.

Schowengerdt, R. A. [1983]. *Techniques for Image Processing and Classification in Remote Sensing*, Academic Press, New York.

Schreiber, W. F. [1956]. "The Measurement of Third Order Probability Distributions of Television Signals." *IRE Trans. Info. Theory*, vol. IT-2, pp. 94 – 105.

Schreiber, W. F. [1967]. "Picture Coding." *Proc. IEEE* (Special issue on Redundancy Reduction), vol. 55, pp. 320 – 330.

Schreiber, W. F., and Knapp, C. F. [1958]. "TV Bandwidth Reduction by Digital Coding." *Proc. IRE National Convention*, pt. 4, pp. 88 – 99.

Schwartz, J. W., and Barker, R. C. [1966]. "Bit-Plane Encoding: A Technique for Source Encoding." *IEEE Trans. Aerosp. Elec. Systems*, vol. AES-2, no. 4, pp. 385 – 392.

Serra, J. [1982]. *Image Analysis and Mathematical Morphology*, Academic Press, New York.

Serra, J. (ed.) [1988]. *Image Analysis and Mathematical Morphology*, vol. 2, Academic Press, New York.

Sezan, M. I., Rabbani, M., and Jones, P. W. [1989]. "Progressive Transmission of Images Using a Prediction/Residual Encoding Approach." *Opt. Eng.*, vol. 28, no. 5, pp. 556 – 564.

Shack, R. V. [1964]. "The Influence of Image Motion and Shutter Operation on the Photographic Transfer Function." *Appl. Opt.*, vol. 3, pp. 1171 – 1181.

Shafarenko, L., Petrou, M., and Kittler, J. [1998]. "Histogram-Based Segmentation in a Perceptually Uniform Color Space." *IEEE Trans. Image Processing*, vol. 7, no. 9, pp. 1354 –

1358.

Shaked, D., and Bruckstein, A. M. [1998]. "Pruning Medial Axes," *Computer Vision and Image Understanding*, vol. 69, no. 2, pp. 156–169.

Shannon, C. E. [1948]. "A Mathematical Theory of Communication." *The Bell Sys. Tech. J.*, vol. XXVII, no. 3, pp. 379–423.

Shapiro, L. G., and Stockman, G. C. [2001]. *Computer Vision*, Prentice Hall, Upper Saddle River, N.J.

Shapiro, V. A. [1996]. "On the Hough Transform of Multi-Level Pictures," *Pattern Recog.*, vol. 29, no. 4, pp. 589–602.

Shariat, H., and Price, K. E. [1990]. "Motion Estimation with More Than Two Frames." *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 5, pp. 417–434.

Sheppard, J. J., Jr., Stratton, R. H., and Gazley, C., Jr. [1969]. "Pseudocolor as a Means of Image Enhancement." *Am. J. Optom. Arch. Am. Acad. Optom.*, vol. 46, pp. 735–754.

Shi, F. Y., and Wong, W.-T. [1994]. "Fully Parallel Thinning with Tolerance to Boundary Noise," *Pattern Recog.*, vol. 27, no. 12, pp. 1677–1695.

Shih, F. Y. C., and Mitchell, O. R. [1989]. "Threshold Decomposition of Gray-Scale Morphology into Binary Morphology." *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 1, pp. 31–42.

Sid-Ahmed, M.A. [1995]. *Image Processing: Theory, Algorithms, and Architectures*, McGraw-Hill, New York.

Sikora, T. [1997]. "MPEG Digital Video-Coding Standards." *IEEE Signal Processing*, vol. 14, no. 5, pp. 82–99.

Simon, J. C. [1986]. *Patterns and Operators: The Foundations of Data Representations*, McGraw-Hill, New York.

Sklansky, J., Chazin, R. L., and Hansen, B. J. [1972]. "Minimum-Perimeter Polygons of Digitized Silhouettes." *IEEE Trans. Comput.*, vol. C-21, no. 3, pp. 260–268.

Smirnov, A. [1999]. *Processing of Multidimensional Signals*, Springer-Verlag, New York.

Smith, A. R. [1978]. "Color Gamut Transform Pairs." *Proc. SIGGRAPH'78*, published as *Computer Graphics*, vol. 12, no. 3, pp. 12–19.

Smith, M. J. T., and Barnwell, T. P. III [1984]. "A Procedure for Building Exact Reconstruction Filter Banks for Subband Coders." *Proc.* IEEE Int. Conf. *Acoust., Speech, and Signal Proc.*, San Diego, Calif.

Smith, M. J. T., and Barnwell, T. P. III [1986]. "Exact Reconstruction Techniques for Tree-Structured Subband Coders." *IEEE Trans. On Acoust., Speech, and Signal Proc.*, vol. 34, no. 3, pp. 434–441.

Sobel, I. E. [1970]. "Camera Models and Machine Perception," Ph.D. dissertation, Stanford University, Palo Alto, Calif.

Sonka, M., Hlavac, V., and Boyle, R. [1999]. *Image Processing, Analysis, and Machine Vi-*

sion, 2nd ed., PWS Publishing, New York.

Stark, H. (ed.) [1987]. *Image Recovery: Theory and Application*, Academic Press, New York.

Stark, J. A. [2000]. "Adaptive Image Contrast Enhancement Using Generalizations of Histogram Equalization," *IEEE Trans. Image Processing*, vol. 9, no. 5, pp. 889–896.

Stockham, T. G., Jr. [1972]. "Image Processing in the Context of a Visual Model." *Proc. IEEE*, vol. 60, no. 7, pp. 828–842.

Storer, J. A., and Reif, J. H., eds. [1991]. *Proceedings of DDC '91*, IEEE Computer Society Press, Los Alamitos, Calif.

Strang, G., and Nguyen, T. [1996]. *Wavelets and Filter Banks*, Wellesley-Cambridge Press Wellesley, Mass.

Stumpff, K. [1939]. *Tafeln und Aufgaben zur Harmonischen Analyse und Periodogrammrechnung*, Julius Springer, Berlin.

Sussner, P., and Ritter, G. X. [1997]. "Decomposition of Gray-Scale Morphological Templates Using the Rank Method," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 6, pp. 649–658.

Swets, D. L., and Weng, J. [1996]. "Using Discriminant Eigenfeatures for Image Retrieval," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, no. 8, pp. 1831–1836.

Symes, P. D. [2001]. *Video Compression Demystified*. McGraw-Hill, New York.

Sze, T. W., and Yang, Y. H. [1981]. "A Simple Contour Matching Algorithm." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-3, no. 6, pp. 676–678.

Tanaka, E. [1995]. "Theoretical Aspects of Syntactic Pattern Recognition," *Pattern Recog.*, vol. 28, no. 7 pp. 1053–1061.

Tanimoto, S. L. [1979]. "Image Transmission with Gross Information First." *Comput. Graphics Image Proc.*, vol. 9, pp. 72–76.

Tasto, M., and Wintz, P. A. [1971]. "Image Coding by Adaptive Block Quantization." *IEEE Trans. Comm. Tech.*, vol. COM-19, pp. 957–972.

Tasto, M., and Wintz, P. A. [1972]. "A Bound on the Rate-Distortion Function and Application to Images." *IEEE Trans. Info. Theory*, vol. IT-18, pp. 150–159.

Teh, C. H., and Chin, R. T. [1989]. "On the Detection of Dominant Points on Digital Curves." *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 8, pp. 859–872.

Thévenaz, P., and Unser, M [2000]. "Optimization of Mutual Information for Multiresolution Image Registration." *IEEE Trans. Image Processing*, vol. 9, no. 12, pp. 2083–2099.

Thomas, L. H. [1963]. "Using a Computer to Solve Problems in Physics." *Application of Digital Computers*, Ginn, Boston.

Thomason, M. G., and Gonzalez, R. C. [1975]. "Syntactic Recognition of Imperfectly Specified Patterns." *IEEE Trans. Comput.*, vol. C-24, no. 1, pp. 93–96.

Thompson, W. B. (ed.) [1989]. "Special Issue on Visual Motion." *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 5, pp. 449–541.

Thompson, W. B., and Barnard, S. T. [1981]. "Lower-Level Estimation and Interpretation of

Visual Motion." *Computer*, vol. 14, no. 8, pp. 20 – 28.

Thorell, L. G., and Smith, W. J. [1990]. *Using Computer Color Effectively*, Prentice Hall, Upper Saddle River, N.J.

Tian, J., and Wells, R. O., Jr. [1995]. *Vanishing Moments and Wavelet Approximation*, Technical Report CML TR-9501, Computational Mathematics Lab., Rice University, Houston, Texas.

Tomita, F., Shirai, Y., and Tsuji, S. [1982]. "Description of Texture by a Structural Analysis." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-4, no. 2, pp. 183 – 191.

Topiwala, P. N. (ed.) [1998]. *Wavelet Image and Video Compression*, Kluwer Academic Publishers, Boston, Mass.

Tou, J. T., and Gonzalez, R. C. [1974]. *Pattern Recognition Principles*, Addison-Wesley, Reading, Mass.

Toussaint, G. T. [1982]. "Computational Geometric Problems in Pattern Recognition." In *Pattern Recognition Theory and Applications*, Kittler, J., Fu, K. S., and Pau, L. F. (eds.), Reidel, New York, pp. 73 – 91.

Tsai, J.-C., Hsieh, C.-H, and Hsu, T.-C. [2000]. "A New Dynamic Finite-State Vector Quantization Algorithm for Image Compression," *IEEE Trans. Image Processing*, vol. 9, no. 11, pp. 1825 – 1836.

Tsujii, O., Freedman, M. T., and Mun, K. S. [1998]. "Anatomic Region-Based Dynamic Range Compression for Chest Radiographs Using Warping Transformation of Correlated Distribution," *IEEE Trans. Medical Imaging*, vol. 17, no. 3, pp. 407 – 418.

Udpikar, V. R., and Raina, J. P. [1987]. "BTC Image Coding Using Vector Quantization." *IEEE Trans. Comm.*, vol. COM-35, no. 3, pp. 352 – 356.

Ueda, N. [2000]. "Optimal Linear Combination of Neural Networks for Improving Classification Performance," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 2, pp. 207 – 215.

Ullman, S. [1981]. "Analysis of Visual Motion by Biological and Computer Systems," *IEEE Computer*, vol. 14, no. 8, pp. 57 – 69.

Umbaugh, S. E. [1998]. *Computer Vision and Image Processing: A Practical Approach Using CVIPtools*, Prentice Hall, Upper Saddle River, N.J.

Umeyama, S. [1988]. "An Eigendecomposition Approach to Weighted Graph Matching Problems." *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, no. 5, pp. 695 – 703.

Unser, M. [1995]. "Texture Classification and Segmentation Using Wavelet Frames." *IEEE Trans. on Image Processing*, vol. 4, no. 11, pp. 1549 – 1560.

Unser, M., Aldroubi, A., and Eden, M. [1993]. "A Family of Polynomial Spline Wavelet Transforms." *Signal Proc.*, vol. 30, no. 2, pp. 141 – 162.

Unser, M., Aldroubi, A., and Eden, M. [1993]. "B-Spline Signal Processing, Parts I and II." *IEEE Trans. Signal Proc.*, vol. 41, no. 2, pp. 821 – 848.

Unser, M., Aldroubi, A., and Eden, M. [1995]. "Enlargement or Reduction of Digital Images with Minimum Loss of Information," *IEEE Trans. Image Processing*, vol. 4, no. 5, pp. 247 –

257.

Vaidyanathan, P. P., and Hoang, P.-Q. [1988]. "Lattice Structures for Optimal Design and Robust Implementaion of Two-Channel Perfect Reconstruction Filter Banks." *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. 36, no. 1, pp. 81 – 94.

Vailaya, A., Jain, A., and Zhang, H. J. [1998]. "On Image Classification: City Images vs. Landscapes," *Pattern Recog.*, vol. 31, no. 12, pp. 1921 – 1935.

Vetterli, M. [1986]. "Filter Banks Allowing Perfect Reconstruction Signal Proc., vol. 10, no. 3, pp. 219 – 244".

Vetterli, M., and Kovacevic, J. [1995]. *Wavelets and Suband Coding*, Prentice Hall, Englewood Cliffs, N.J.

Voss, K., and Suesse, H. [1997]. "Invariant Fitting of Planar Objects by Primitives," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 1, pp. 80 – 84.

Vuylsteke, P., and Kittler, J. [1990]. "Edge-Labeling Using Dictionary-Based Relaxation." *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 2, pp. 165 – 181.

Walsh, J. W. T. [1958]. *Photometry*, Dover, New York.

Wang, G., Zhang, J., and Pan, G.-W. [1995]. "Solution of Inverse Problems in Image Processing by Wavelet Expansion." *IEEE Trans. Image Processing*, vol. 4, no. 5, pp. 579 – 593.

Wang, Y.-P., Lee, S. L., and Toraichi, K. [1999]. "Multiscale Curvature-Based Shape Representation Using β -Spline Wavelets." *IEEE Trans. Image Processing*, vol. 8, no. 11, pp. 1586 – 1592.

Wang, Z., Rao, K. R., and Ben-Arie, J. [1996]. "Optimal Ramp Edge Detection Using Expansion Matching," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, no. 11, pp. 1092 – 1097.

Watt, A. [1993]. *3D Computer Graphics*, 2nd ed., Addison-Wesley, Reading, Mass.

Wechsler [1980]. "Texture Analysis—A Survey." *Signal Process*, vol. 2, pp. 271 – 280.

Wei, D., Tian, J., Wells, R. O., Jr., and Burrus, C. S. [1998]. "A New Class of Biorthogonal Wavelet Systems for Image Transform Coding." *IEEE Trans. Image Processing*, vol. 7, no. 7, pp. 1000 – 1013.

Weinberger, M. J., Seroussi, G., and Sapiro, G. [2000]. "The LOCO-I Lossless Image Compression Algorithm: Principles and Standardization into JPEG-LS," *IEEE Trans. Image Processing*, vol. 9, no. 8, pp. 1309 – 1324.

Westenberg, M. A., and Roerdink, J. B. T. M. [2000]. "Frequency Domain Volume Rendering by the Wavelet X-ray Transform." *IEEE Trans. Image Processing*, vol. 9, no. 7, pp. 1249 – 1261.

Weszka, J. S. [1978]. "A Survey of Threshold Selection Techniques." *Comput. Graphics Image Proc.*, vol. 7, pp. 259 – 265.

White, J. M., and Rohrer, G. D. [1983]. "Image Thresholding for Optical Character Recognition and Other Applications Requiring Character Image Extraction." *IBM J. Res. Devel.*, vol. 27,

no. 4, pp. 400 - 411.

Widrow, B. [1962]. "Generalization and Information Storage in Networks of 'Adaline' Neurons." In *Self-Organizing Systems 1962*, Yovitz, M. C., et al. (eds.), Spartan, Wash-ington, D. C., pp. 435 - 461.

Widrow, B., and Hoff, M. E. [1960]. "Adaptive Switching Circuits." *1960 IRE WESCON Convention Record*, Part 4, pp. 96 - 104.

Widrow, B., and Stearns, S. D. [1985]. *Adaptive Signal Processing*, Prentice Hall, Englewood Cliffs, N.J.

Wiener, N. [1942]. *Extrapolation, Interpolation, and Smoothing of Stationary Time Se-ries*, the MIT Press, Cambridge, Mass.

Wilburn, J. B. [1998]. "Developments in Generalized Ranked-Order Filters," *J. Opt. Soc. Amer.-A. Optics, Image Science, and Vision*, vol. 15, no. 5, pp. 1084 - 1099.

Windyga, P. S. [2001]. "Fast Impulsive Noise Removal," *IEEE Trans. Image Processing*, vol. 10, no. 1, pp. 173 - 179.

Wintz, P. A. [1972]. "Transform Picture Coding." *Proc. IEEE*, vol. 60, no. 7, pp. 809 - 820.

Witten, I. H., Neal, R. M., and Cleary, J. G. [1987]. "Arithmetic Coding for Data Compression." *Comm. ACM*, vol. 30, no. 6, pp. 520 - 540.

Wolff, R. S., and Yaeger, L. [1993]. *Visualization of Natural Phenomena*, Springer-Verlag, New York.

Woods, J. W., and O'Neil, S. D. [1986]. "Subband Coding of Images." *IEEE Trans. Acous. Speech Signal Proc.*, vol. ASSP-35, no. 5, pp. 1278 - 1288.

Woods, R. E., and Gonzalez, R. C. [1981]. "Real-Time Digital Image Enhancement." *Proc. IEEE*, vol. 69, no. 5, pp. 643 - 654.

Xu, Y., Weaver, J. B., Healy, D. M., Jr., and Lu, J. [1994]. "Wavelet Transform Domain Fil-ters: A Spatially Selective Noise Filtration Technique." *IEEE Trans. Image Processing*, vol. 3, no. 6, pp. 747 - 758.

Yachida, M. [1983]. "Determining Velocity Maps by Spatio-Temporal Neighborhoods from Image Sequences." *Comput. Vis. Graph. Image Proc.*, vol. 21, no. 2, pp. 262 - 279.

Yamazaki, Y., Wakahara, Y., and Teramura, H. [1976]. "Digital Facsimile Equipment 'Quick-FAX' Using a New Redundancy Reduction Technique." *NTC '76*, pp. 6.2-1 - 6.2-5.

Yang, X., and Ramchandran, K. [2000]. "Scalable Wavelet Video Coding Using Aliasing Reduced Hierarchical Motion Compensation," *IEEE Trans. Image Processing*, vol. 9, no. 5, pp. 778 - 791.

Yates, F. [1937]. "The Design and Analysis of Factorial Experiments." Commonwealth Agricultural Bureaux, Farnam Royal, Burks, England.

Yitzhaky, Y., Lantzman, A., and Kopeika, N. S. [1998]. "Direct Method for Restoration of Motion Blurred Images," *J. Opt. Soc. Amer.-A. Optics, Image Science, and Vision*, vol. 15, no. 6, pp. 1512 - 1519.

You, J., and Bhattacharya, P. [2000]. "A Wavelet-Based Coarse-to-Fine Image Matching Scheme in a Parallel Virtual Machine Environment." *IEEE Trans. Image Processing*, vol. 9, no. 9, pp. 1547 – 1559.

Yu, D., and Yan, H. [2001]. "Reconstruction of Broken Handwritten Digits Based on Structural Morphology," *Pattern Recog.*, vol. 34, no. 2, pp. 235 – 254.

Yu, S. S., and Tsai, W. H. [1990]. "A New Thinning Algorithm for Gray-Scale Images." *Pattern Recog.*, vol. 23, no. 10, pp. 1067 – 1076.

Yuan, M., and Li, J. [1987]. "A Production System for LSI Chip Anatomizing." *Pattern Recog. Letters*, vol. 5, no. 3, pp. 227 – 232.

Zahn, C. T., and Roskies, R. Z. [1972]. "Fourier Descriptors for Plane Closed Curves." *IEEE Trans. Comput.*, vol. C-21, no. 3, pp. 269 – 281.

Zhang, T. Y., and Suen, C. Y. [1984]. "A Fast Parallel Algorithm for Thinning Digital Patterns." *Comm. ACM*, vol. 27, no. 3, pp. 236 – 239.

Zhu, H., Chan F. H. Y., and Lam, F. K. [1999]. "Image Contrast Enhancement by Constrained Local Histogram Equalization," *Computer Vision and Image Understanding*, vol. 73, no. 2, pp. 281 – 290.

Zhu, P., and Chirlian, P. M. [1995]. "On Critical Point Detection of Digital Shapes," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, no. 8, pp. 737 – 748.

Zimmer, Y., Tepper, R., and Akselrod, S. [1997]. "An Improved Method to Compute the Convex Hull of a Shape in a Binary Image," *Pattern Recog.*, vol. 30, no. 3, pp. 397 – 402.

Ziou, D. [2001]. "The Influence of Edge Direction on the Estimation of Edge Contrast and Orientation," *Pattern Recog.*, vol. 34, no. 4, pp. 855 – 863.

Ziv, J., and Lempel, A. [1977]. "A Universal Algorithm for Sequential Data Compression." *IEEE Trans. Info. Theory*, vol. IT-23, no. 3, pp. 337 – 343.

Ziv, J., and Lempel, A. [1978]. "Compression of Individual Sequences Via Variable-Rate Coding." *IEEE Trans. Info. Theory*, vol. IT-24, no. 5, pp. 530 – 536.

Zucker, S. W. [1976]. "Region Growing: Childhood and Adolescence." *Comput. Graphics Image Proc.*, vol. 5, pp. 382 – 399.

Zugaj, D., and Lattuati, V. [1998]. "A New Approach of Color Images Segmentation Based on Fusing Region and Edge Segmentation Outputs," *Pattern Recog.*, vol. 31, no. 2, pp. 105 – 113.